# Journal of Practical Ethics

# CONTENTS

The Journal of Practical Ethics is available online, free of charge, at:
http://jpe.ox.ac.uk

*Editorial Policy*
The *Journal of Practical Ethics* is an invitation only, blind-peer-reviewed journal. It is entirely open access online, and print copies may be ordered at cost price via a print-on-demand service. Authors and reviewers are of-fered an honorarium for accepted articles. The journal aims to bring the best in academic moral and political philosophy, applied to practical mat-ters, to a broader student or interested public audience. It seeks to promote informed, rational debate, and is not tied to any one particular viewpoint. The journal will present a range of views and conclusions within the ana-lytic philosophy tradition. It is funded through the generous support of the *Uehiro Foundation in Ethics and Education.*

# The Neglected Harms of Beauty: Beyond Engaging Individuals

HEATHER WIDDOWS

*University of Birmingham*

## ABSTRACT

This paper explores the neglected 'harms-to-others' which result from increased attention to beauty, increased engagement in beauty practices and rising minimal beauty standards. In the first half of the paper I consider the dominant discourse of beauty harms – that of ethics and policy – and argue that this discourse has over-focused on the agency of, and possible harms to, recipients of beauty practices. I introduce the feminist discourse which recognises a general harm to all women and points towards an alternative understanding; although it too focuses on engaging individuals. I argue over-focusing on harms to engaging individuals is somewhat surprising especially in liberal contexts, as this harm can broadly be regarded as 'self-harm' (done by individuals to themselves, or by others employed by individuals to do so). The focus on engaging individuals has resulted in the neglect of significant and pressing harms-to-others in theory, policy and practice. In the second half of the paper I turn to actual and emerging harms-to-others. I focus on three particular harms-to-others as examples of the breadth and depth of beauty harms: first, direct harm to providers; second, indirect but specific harm to those who are 'abnormal'; and third, indirect and general harm to all. I conclude that, contrary to current discourses, harms-to-others need to be taken into account to avoid biased and partial theorising and counter-productive policy-making. I advocate recasting beauty, in a parallel way to smoking, as a matter of public health rather than individual choice.

# 1. INTRODUCTION

In this paper I explore the neglected harms of beauty. Some potential harms of beauty are widely recognised as issues of concern and are matters of considerable debate but others are almost wholly neglected. Primarily policy-makers, ethicists and lawyers have focused on harms to engaging individuals; cosmetic surgery recipients and beauty practices users. I argue that harms arising from such practices can largely be classified as 'self-harm'; assuming a very broad definition of self-harm is adopted. A broad definition of self-harm includes the practices that individuals do to themselves, and those which they employ others to do to them. On this definition the focus on engaging individuals rather than on harms-to-others is perplexing as, on standard liberal models, harms-to-others are considered more morally significant than harms to the self. Prohibiting self-harm is often considered troublingly paternalistic in liberal contexts and yet in the beauty debate attention has almost exclusively been on the harms to engaging individuals. My aim is to highlight and investigate this striking gap and to argue that this neglect is a significant failure of theory and policy. I argue that the increased valuing of appearance in a visual and virtual culture, coupled with rising engagement and a rise in what is required to meet minimal standards of beauty, is already harmful to others.[1] Moreover, if current trends continue and appearance continues to matter more then harms which are currently nascent are likely to become prevalent. I suggest that beauty harms should be reframed as public health concerns.

In order to make this argument, in section 2, I outline two discourses in which the harms of beauty are currently discussed; those of ethics and policy and feminist political philosophy. I do this to justify my claim that such debates have primarily focused on engaging individuals; and to consider the connection the feminist debate makes between beauty practices and a collective harm to all women. In section 3 I turn to overt harms-to-others; to harms beyond those which fall on currently en-

---

1.    In the paper I assume that minimal beauty standards are rising. This is not a claim that all standards are everywhere the same, or that there have not been more demanding beauty standards historically, undoubtedly there have; footbinding being a classic example. Rather it is a claim that as a global beauty ideal of thinness, firmness, smoothness and youth, emerges so globally minimal standards rise and are standardised and normalised (Widdows, forthcoming 2018).

gaging individuals. I introduce a number of possible harms-to-others to indicate the breadth of potential beauty harms. In section 4 I consider three harms-to-others. Each of these harms is an example of a different type of beauty harm. I have selected these three harms in order to illustrate the range and extent of potential beauty harms and to show that in some cases clear harms are already being ignored and likely emerging harms are not being recognised and anticipated. Accordingly there are harms already occurring or emerging which merit response. The three harms I explore are: First, the direct harm to providers of beauty practices; second, the indirect but specific harm to those who are 'abnormal' and can never conform to minimal standards of beauty; and third, the indirect and general harm to all. I conclude that, contrary to current discourses, the most widespread and significant current and likely harms of beauty are not harms to engaging individuals, but to others. Accordingly theoretical work, as well as policy and practice recommendations, must pay attention to harms-to-others, as well as to engaging individuals, if they are to avoid advocating biased or counter-productive conceptions, policies and practices.

## 2. STANDARD BEAUTY DISCOURSES

There are two separate discourses which have focused on the harms of beauty practices. The first is that of policy-makers, ethicists and lawyers, who have focused on the harms and risks to those who engage in beauty practices. The second is feminist critiques of beauty practices as instances of gender injustice. The only exception to the dominant focus on the harms to engaging individuals has been in the ethics of non-human animals, where the testing of beauty products is an issue of concern.

*2.1 Safety, information and consent*

The first discourse is primarily about safety, governance, and professional duties with regard to invasive procedures. There is concern that recipients are electing to have procedures or engage in beauty practices which are unduly risky; carried out by practitioners who lack extensive training in the procedures they are administering in unregulated or under-regulated premises, using untested products and without adequate information or sufficiently robust consent procedures. Such worries are heightened by scandals and rising public concern. For instance, the Poly Implant Prosthese (PIP) scandal of 2010 arose when breast implants were made from silicone intended

for industrial rather than medical purposes (Keogh, 2012). Similar scandals attach to non-surgical processes and there are constant horror stories about injectables and fillers; and complications include infection, lumpiness, blood vessel blockages, tissue death, allergic reactions, prolonged swelling and bruising and even blindness (Keogh, 2013, 24). At the most routine end of the spectrum, there are worries about practices such as skin lightening and tanning; for example, the World Health Organisation considers skin lightening a public health risk across parts of Africa, Asia and Latin America (WHO, 2011).

Such scandals have led to numerous reports and recommendations, with particular attention being paid to the most risky practices, those of surgery and injectables. Most prominent in the UK is the Department of Health's 'Review of the Regulation of Cosmetic Interventions' chaired by Bruce Keogh; often referred to just as 'the Keogh review' (Keogh, 2013). The Keogh review recommended changes in three areas: first, in provision, calling for safe products and skilled and responsible providers; second, in care, to ensure an informed and protected public; and third, in redress. While there has not been the legislative response Keogh recommended professional bodies have taken action. For example, the General Medical Council introduced 'Guidance for Doctors who offer Cosmetic Interventions' which emphasized adequate training and experience, ensuring realistic expectations, responsible marketing and the importance of doctors consenting patients in person (GMC, 2016). This was supplemented by The Royal College of Surgeons' publication, 'Professional standards for Cosmetic Surgery', which provides detailed instructions on how to meet professional standards (RCS, 2016). In addition, in 2017 the Royal College of Surgeons launched a certification scheme for appropriately trained and experienced surgeons.[2] The goals of such initiatives are to improve the safety of procedures and to ensure that potential recipients are fully informed and that best practice consent procedures are implemented across the board. In particular, that recipients are fully informed of risks and possible complications overtime; that there are no financial inducements to pressure recipients into quick decisions; and that there is a no-penalty cooling off period.

Such responses only go so far. They do not address concerns about surgeons who do not fall under the auspices of UK professional bodies: recipients travel abroad to access cheaper procedures and surgeons are flown in by cosmetic companies to do

2.   Members of the public can search a specialist register; however while some cosmetic surgery providers only employ registered surgeons non-registered surgeons continue to practice.

multiple operations.[3] Nor do they address the lack of regulation, governance and best practice of non-surgical procedures. For instance, while Botox requires a prescription in the UK, it can be bought from the internet, and dermal fillers are almost wholly unregulated. The Keogh review memorably puts it that "a person having a non-surgical cosmetic intervention has no more protection and redress than someone buying a ballpoint pen or a toothbrush" (Keogh, 2013, 5). The review continues:

> *"Dermal fillers are a particular cause for concern as anyone can set themselves up as a practitioner, with no requirement for knowledge, training or previous experience. Nor are there sufficient checks in place with regard to product quality – most dermal fillers have no more controls than a bottle of floor cleaner" (Keogh, 2013, 5).*

In terms of harms to recipients the regulation of non-surgical treatments is a pressing area for policy makers to address. Non-surgical procedures are far less regulated than surgical procedures and yet it is here where uptake is rising exponentially.[4] Although figures should be used cautiously, that these procedures are increasingly commonplace is not contested; accurate data is notoriously lacking across beauty practices and there is no systematic recording of the numbers of surgical and non-surgical procedures, outcomes, complications and side-effects.[5] Part of ensuring that products and premises are safe and the practitioners are well-trained is about protecting engaging individuals from harm-from-others. Serious physical and psychological harm can result from both surgical and non-surgical procedures.[6] Some harms are

3.   So called 'fly-in-fly-out' or 'seagull' surgeons are an increasing phenomenon and the primary reason for cosmetic surgery tourism is to attain cheaper treatments (Griffiths and Mullock, 2017).

4.   There is no centralised reporting of the number of such procedures and therefore estimates of the numbers of procedures carried out in the UK or globally are impossible to attain. However, that there is a rise in Botox and other non-surgical but invasive procedures (such as fillers, non-surgical face lifts and chemical peels) is undoubtedly the case. All surveys, however incomplete, suggest that Botox rises year on year. For instance, the 2016 statistics for procedures carried out in the US describes Botox as the most popular non-surgical procedure with over 4.5 million procedures carried out by registered medics in 2016 (ASAPS, 2016, 7).

5.   The lack of data, particularly extensive and robust data, is a constant complaint of policy-makers in the beauty context. The frustration is evident in the Keogh review which comments that, "there is no central collection of data on the complications following cosmetic interventions and hence no information on the type or frequency of complications" (Keogh, 2013, 39). This frustration is echoed in the Nuffield Council on Bioethics Report, *The Ethics of Cosmetic Procedures* (Nuffield, 2017).

6.   Physical harms include standard harms of surgery (including adverse reactions to anaesthesia, bleeding, scaring and risks of infection and complications) as well as added risks which attach to using implants (either from additional risks of infection and rejection or from harms from the sometimes toxic implants, especially when they rupture or leak). Psychological harms are more contested and some argue offset by psychological benefit, but possible psychological harms include increased attention to appearance leading to increased body image anxiety and its negative consequences.

foreseeable and intentional; for instance, that the recipients of PIP implants would be harmed was wholly foreseeable. This was a direct harm, and one the women did not do to themselves and which could not be classed as self-harm.

Protecting individuals from using unsafe products is part of protecting individuals from harm-from-others. Moreover, professional bodies are, by definition, concerned with regulating the profession, and the emphasis on safe products and procedures serves to protect the profession and the practitioners within it as well as engaging individuals. Yet, assuming that providers are acting ethically, providing services and products in good faith and informing recipients of the risks, then both practitioners and recipients are protected, and harms to engaging individuals are effectively self-harm. For example, in cases where surgery goes wrong but no one has acted wrongly harms which arise are best understood as 'self-harm'. If individuals are made aware of the risks—and there are always risks—and have fully consented then they are taking responsibility for bad outcomes which they know might occur. All surgery can have complications, even when performed by the best surgeon in the best setting. That recipients do take responsibility for bad outcomes when no one is at fault is borne out by the testimonies of women who engage.[7] Whether or not this is ethical, and whether these decisions are really as free as the consent model assumes is something I debate elsewhere (Widdows, forthcoming 2018). However, on the liberal model where individuals are permitted to engage in harmful and risky practices as long as they are fully informed, weigh the risks and benefits, and formally consent such harms are, rightly or wrongly, placed at the door of the consenting individual. Accordingly harms which befall engaging individuals can be classified in some very broad sense as 'self-harm'.

If such harm is self-harm, then the lack of attention paid to harms-to-others is perplexing. One possible explanation for focusing only on those engaging rather than others is the difficulty of tracking harm-to-others. Collecting comprehensive data on the harms to those who engage in practices may be challenging, and currently lacking, but collecting data on those who do not engage is more challenging. This is particularly true when we consider indirect harm, where there is no causal link between an action and a harmful outcome. While direct harm to others (such as that to providers which will be considered in section 4.1) is possible to evidence in a similar way to direct harm to recipients, indirect or communal harms, for instance,

---

7.    For example 'Betty', a cosmetic surgery recipient, voices this, when she says that having had surgery "you have to stand behind the decision" (Davis, 1995, 145).

harms resulting from changing social norms and expectations, is far more difficult to track and document. Evidence of general or communal harms is never causal, but requires the tracking of patterns, trends and correlations and making informed, but always interpretive, deductions from data. Given this, claims of general, communal and group harm are harder to substantiate and are not definitive in the way which causal harm to engaging individuals is; although the harms are directly experienced by individuals in that it is individuals who feel the pressure to conform to more demanding beauty ideals and social norms. Evidencing this harm is far more complex than evidencing the harms of physical disfiguration or pain. The operation or procedure is the cause of physical disfigurement or pain; by contrast what is the cause of increased body-dissatisfaction? To be sure feeling increased pressure to engage in more beauty practices correlates with the increased use and normalisation of practices, but the exact nature of the relationship is indeterminate. Therefore, despite the significant damage done to very many, it may be that the lack of direct causal evidence is sufficient for policy makers to shy away from recognising and addressing such harms.

### 2.2 Inequality, inferiority and gendered harm

The second discourse about the harms of beauty practices is found in feminist political philosophy. At first glance this discourse appears to be wholly distinct from the discourse in medical ethics, law and policy. However, on investigation, it too tends to focus on the harms to women who engage but, and importantly for this paper, the feminist discourse also highlights communal or general harm to all woman.

This type of argument originates with radical second-wave feminist thinkers such as Andrea Dworkin. Dworkin introduces footbinding as an extreme but typical beauty practice. It serves to sexualise, physically constrain and physically hurt women; it establishes women "as ornaments, as sexual playthings, as sexual constructs" (Dworkin, 1974, 106). Dworkin argues that beauty practices in general serve this function: they make women 'women', by showing women are different from and subservient to men. Moreover, that they are painful is not accidental, but "teaches an important lesson: no price is too great, no process too repulsive, no operation too painful for the women who would be beautiful" (Dworkin, 1974, 115). The use of beauty practices to make women 'women' and to transform them into stereotypically sexual objects, such feminists argue, is harmful and costly for individual women, and harmful to all women.

Clare Chambers makes a similar argument, but within a liberal framework and using language which is more familiar to current debates. She argues that engagement in beauty practices, such as elective breast augmentation, is harmful because it compromises the core liberal value of equality. For Chambers, individual women are harmed as a result of conforming to patriarchal norms. While an individual woman's choice to engage may be free from the desires of any actual man, it is not, and cannot be, free from the constraint of patriarchal norms. To illustrate, she compares breast implants and knee implants highlighting their cultural meaning:

*"Why on earth would anyone want to have surgery to insert heavy and dangerous alien objects into her body if there were not social meaning to, or social payoff from, the practice? A woman who did want to have breast implants in such a society would be like someone who wanted to have cosmetic knee implants" (Chambers, 2008, 40).*

Chambers argues that it is rational for women to choose to engage in beauty practices, but only because of the unjust context. In her analysis individual women are harmed by engaging in risky, costly and painful practices in order to conform to discriminatory and harmful social norms. She questions whether even the most informed of consents could be truly agentic as "gender inequality is so deeply entrenched in social norms that individual free choice cannot overcome it" (Chambers, 2008, 8).[8] However, for this paper what is important to note is that the feminist debate—like the policy and ethics debate—focuses on harms to engaging individuals. The ethics, law and policy debate seeks to limit harm to engaging individuals by enhancing agency. For example, the reason emphasis is placed on improving consent processes because it is assumed that individuals who have information can make autonomous and robust decisions. By contrast, the feminist debate is concerned that such decisions, which benefit men rather than women, can never be free (Jeffreys, 2005, 32). Despite these differences the focus for both is on engaging individuals.

However, the feminist debate, while focusing on engaging individuals, high-

8.    Jessica Laimann is more optimistic about informed consent from a feminist perspective and suggests that: "In addition to education campaigns that raise general awareness of the discriminatory and objectifying nature of the practice of breast implants, the relevant information could be specifically communicated as part of the physician-patient-consultation, or on the medical consent form that patients are required to sign before breast implant surgery. The relevant information would not only include reference to the status harm that having breast implants might entail, but could also feature information about the harm effects breast implants might have on other women by increasing acceptance and influence of the relevant norm" (Laimann, 2015, 56).

lights one important harm-to-others; the general harm or communal harm which all women suffer. To return to Chambers, she argues that "the problem with disciplinary appearance norms is not just that they are different for men and women, and not just that they are more exacting and expensive (in both time and money) for women, but that their effect is to cast women as inferior" (Chambers, 2008, 29). This then is a general, communal or group harm which falls on all women; all women are being marked as inferior and unequal. It is a status harm. For Chambers, on liberal grounds, States should intervene as "liberal institutions ought to ensure that, wherever possible, pressures to make disadvantageous choices should not fall on a specific group or groups" (Chambers, 2008, 130). Her claim is that choices which are systematically made by one group and not another suggest disadvantage, and choices which are responses to identifiable pressures suggest undue influence. Both systematic influence and/or disadvantage are indications of injustice.[9] In Chambers argument, as in Dworkin's before her, it matters that the requirements of beauty are unequal; that they fall on women and not men. As Sandra Bartky memorably puts it, "soap and water, a shave and routine attention to hygiene may be enough for *him*; for *her* they are not" (Bartky, 1990, 71). The asymmetry or inequality marks women's status as inferior and unequal and permits gendered exploitation and/or subordination.

What I wish to take from this debate is not the gendered nature of beauty harms—an argument I critique elsewhere[10]—but rather the recognition that there might be general and communal harms to all which attach to beauty practices and norms. The status harm falls on all women; irrespective of whether they engage or not. This claim, that social communal norms can be harmful—and to many, perhaps all—is a key argument, and a model upon which I draw. Social norms, in the feminist discourse, can harm: they can impose limitations on what it is possible for individuals to be and do. I will return to communal harms of this broad type in section 4.3.

---

9.    Chambers identifies two key indicators of injustice, the 'disadvantage factor' (choosers are harmed compared to those who choose differently) and the 'influence factor' (identifiable pressures on choosers from the group who choose differently and benefit) (Chambers, 2008, 120).

10.    I argue at length that while there are significant and troubling gendered harms of beauty, for instance from hyper-sexualised norms, it is less possible to claim that *engagement* continues to be a gendered harm as the minimal demands of beauty become increasingly burdensome for men (Widdows, 2017).

## 3. NEGLECTED HARMS-TO-OTHERS

In the last section I claimed that the dominant discourse about the ethics of beauty practices is overwhelmingly focused on and concerned with harms to engaging individuals. This is also the case for the second discourse of feminist political philosophy, although some attention is also paid to the communal and general harm to women as a group. That the medical ethics discourse is so focused is perhaps not surprising. The concern of medical professionals is not to question wider social norms but to do their best for their particular patient within current frameworks. Nor is it surprising that they assume, despite the numerous critiques, that informed consent protects from harm, as this is a standard assumption in medical practice.[11] Yet policy-makers cannot appeal to the same defence. It is exactly the task of policy-makers to consider harms across their jurisdictions and to recognise the interaction of practices, policies and norms. The task of policy-makers, considered broadly, is to put in place governance frameworks which leave individuals space to live the lives they choose as long as they do not harm or unduly proscribe the freedom of others. Some, even liberal, models go further and argue that policy-makers should seek to provide equality of opportunity or conditions of human flourishing for those in their jurisdictions. But, endorsing these stronger claims is not necessary to claim that policy-makers should regulate to protect others from harm and/or prevent the undue restriction of others' freedom. It is exactly the harms-to-others and restriction of others' freedom deriving from increased attention to and engagement with beauty practices which is currently neglected.

While harms to individuals who engage are important and by no means yet fully addressed the harms of beauty are not limited to those which beset engaging individuals. To accurately consider the harms of beauty a broader frame is required, one which can recognise the harms-to-others which result from increasing engagement with beauty practices and the knock on effects of such engagement as engagement is normalised and minimal standards rise. Such harms fall not only on those who choose to engage in beauty practices but on those who do not engage, or who only engage enough to meet minimal standards of beauty, to be 'good enough'.[12] There

---

11.    For instance, Neil Manson and Onora O'Neill argue that consent should not be regarded as protecting autonomy, but rather as a means to waive certain rights (Manson and O'Neill, 2007).

12.    For discussion on the ethical similarity and dissimilarity between the explanatory and justifying narratives for *engagement* of 'to be normal', 'to be good enough', 'to be better' and 'to be perfect', see last section of chapter 5, "Perfectly Normal", of *Perfect Me* (Widdows, forthcoming 2018).

are numerous possible —direct, indirect, individual, group, communal and general— harms which might attach to the rising demands of beauty.[13] In section 4 I will focus on three distinct, actual or likely, harms-to-others which follow from increased engagement with beauty practices. However, these harms are by no means exhaustive of harms-to-others which attach to beauty practices considered broadly, but rather they have been selected to exemplify different types of harm and so show the range and extent of possible beauty harms. Other harms-to-others I could have considered are very general harms of discrimination, harms of unequal distribution of beauty, intergenerational harms, gendered harms from hyper-sexualised beauty norms, and harms to particular racial or ethnic groups. Before I detail three specific harms, let me briefly comment on discrimination and the potential distributive justice harms, as these are increasingly discussed in certain quarters and are indicative of the breadth of possible beauty harms.

Appearance discrimination, 'lookism', has been increasingly discussed and has been compared with sexism and racism (Etcoff, 1999; Swami and Furnham, 2008). Appearance discrimination is a broad category which encompasses a number of possible harms. For example, in some forms it could be considered a general communal harm in that a society which discriminates on appearance grounds creates a toxic environment in which appearance matters more than other goods (a parallel claim is made in normalisation arguments).[14] Alternatively appearance discrimination can be considered group harm; limited to a specific group which is singled out and discriminated against on appearance grounds. Finally—and this is the focus of much of the emerging literature—appearance discrimination can be an individual harm; experienced by individuals who are denied employment or other goods on appearance grounds.

When it comes to individual harm the evidence is contested with regard to the extent to which appearance impacts upon employment and other opportunities. Evidence suggests that there is a small but clear advantage to good looks and a small but clear disadvantage to being classed as unattractive; and such advantages and

---

13.    In this paper I do not distinguish in detail between different types of general, communal and group harms. The reason for this is that my aim is to show that harms beyond engaging individuals must be recognised, hence the purpose is to show that all of these non-individual harms are important, rather than to distinguish between them.

14.    This argument underpins the claims in sections 4.3 that rising minimal standards impact on all. For a detailed account of the process of normalisation see chapter 5 of *Perfect Me*, "Perfectly Normal" (Widdows, forthcoming 2018)

disadvantages apply across domains.[15] Yet some argue that such advantages are over-estimated, and are outweighed by the harms which attach to over-valuing appearance: particularly with regard to self-esteem (which has significant knock on effects for confidence and activity in other areas).[16] Given the current lack of consensus, laws to prevent lookism may be premature: the evidence is contested and legislation might further embed a view that appearance is important and be counter-productive. Moreover, some claim that anti-discrimination laws would be unworkable as it would be difficult to determine those who could be classed as 'unattractive' and thus be subject to such discrimination.[17] In addition, once appearance discrimination was illegal, employers would be unlikely to give appearance as a reason for not appointing, making appearance discrimination particularly hard to prove.

Nonetheless if appearance continues to become ever more valuable and valued in a visual and virtual culture then calls to extend discrimination laws to appearance may increase. Moreover even if legislation on discrimination grounds is not appropriate other beauty harms may require mitigation or redress. For instance, if beauty is regarded as an important good—relevantly similar to health or education—then the distribution of beauty, opportunities to access beauty, or compensation for a lack of beauty, are issues of distributive justice. Beauty is already functioning as such a good in some contexts. For example, women report having surgery in order to ensure continued employment, and professional longevity is increasingly a reason for beauty engagement (Gimlin, 2012). Moreover, some go as far as to argue for a 'right to beauty'; a dominant discourse in Brazil where cosmetic surgery, *plástica*, is widely available in public hospitals (Edmonds, 2007; 2010). In this context Ivo Pitanguy, a famous plastic surgeon, has asserted that "the poor have a right to be beautiful too" (Edmonds, 2010,

15.    For instance, Daniel Hamermesh, draws on numerous studies to show that when it comes to earnings there is a 'beauty premium' and an 'ugly penalty' of approximately 15% (Hamermesh, 2011, 46). Similarly in other domains; the more attractive are more likely to marry (Etcoff, 2011, 65) and to be assumed to have positive personality traits, such as friendliness, competence and intelligence (Eagly et al., 1991).

16.    For example, those who are considered beautiful may experience anxiety and insecurity as a result of heavy investment in their looks (Vantarian, 2009).

17.    Most problematic is who would be protected. William Corbett suggests that because there is not a clearly identified group deciding who would be covered by such legislation would likely dominate any litigation (Corbett, 2007). Moreover, Daniel Hamermesh points out that obesity discrimination cases have been brought on the grounds that obesity is a disability (Hamermesh, 2011, 155). For obvious reasons bringing cases on disability rather than lookist grounds might be preferable for plaintiffs.

14).[18] While the argument in Brazil for a 'right to beauty' is very contextual and connected to wider issues (not least as Brazil's position as a hub for training many of the world's cosmetic surgeons (Edmonds, 2010, 93)), if it is the case that beauty is a significant enabling good, then a lack of beauty, or a lack of access to beauty, is an instance of injustice.

I raise the harms of discrimination and unjust distribution of goods or opportunity, not to draw conclusions (not least as the extent of the harms attached are contested and indeterminate) but to show the possible extent of the harms that might attach to beauty. The harms I wish to focus on in detail in the next section are less contested and more determinant. The harms have been selected to illustrate different types of harm which might attach to beauty; to suggest that different harms fall on different groups, in different ways with various impact; and to show the possible extent of such harms. Thus I seek to highlight both the breadth and depth of potential beauty harms. To this end, I focus on one direct and two indirect harms which are already occurring, emerging and probable. The direct harm is to easily identifiable others; the indirect harm is in one case to identifiable others and in the other it is a general or communal harm which potentially affects all.

## 4. THREE BEAUTY HARMS-TO-OTHERS

### *4.1 Direct harm to providers*

As discussed in section 2 the ethics of beauty practices has largely been concerned with the medical or pseudo-medical practices of cosmetic surgery and the duties of medical professionals. But, while there has been significant attention to the harms to recipients, there has been very little attention on harms to providers. Given that this is a direct harm, where evidence and causal data should not be difficult to find, this omission is particularly glaring.

Harms to providers tend to be in non-medical settings. Very many beauty practices (including invasive practices) are not carried out in medical settings or by

---

18.    Arguments for a 'right to beauty' in Brazil are taking place in a very distinct discourse in which surgery is regarded as treating the mind and addressing multiple needs, and where surgeons believe that "plástica is a form of 'public health' to which the poor should have access" (Edmonds, 2007, 367). Given the 'right to beauty' is being asserted in a context where human rights are generally hard to access and there is little social mobility it may be that this discourse serves other purposes and provides "a popular form of hope" (Edmonds, 2007, 378) or an "imaginary vehicle of ascent" (Edmonds, 2010, 20) when other opportunities are lacking.

medical professionals. If we think, as I argue we should, that cosmetic surgery is better conceived of as a beauty practice rather than a medical practice then we can posit a continuum of beauty practices. At one end of the continuum is minimal grooming (practices such as hair-styling and the routine and often daily application of lotions and potions) at the other end is the most risky type of beauty practice, that of surgery (recognising that surgery too is on a continuum; some is routine, frequent, with quick recovery times and low complication rates and some is complex, novel, has long recovery times and high risks of complication).[19] Conceptualised in this way there is no clear line between procedures which are routine beauty practices, invasive or surgical, nor is there a clear line between who carries out what practices in what settings.[20] The middle of the continuum is particularly opaque and very many beauty practices are not provided by highly trained, privileged and protected medical professionals. Even when they are carried out by medical professionals the professional in question may not have training in the specific procedure.[21] Some beauty practices carry little risk, either to the recipient or to the provider, and others are highly risky to either the recipient or the provider. However while there is significant work on the harms to recipients there is little on the risks to providers of procedures.[22]

The first direct harm to others is to providers who work largely in non-medical settings at the less-invasive end of the beauty practice continuum; hairdressers, nail technicians and beauticians. Beauty practitioners are not highly skilled, or at least their skill-base is not regarded as 'expert' in a way which is equivalent to or approximates with the medical professionals' skill base; although in many instances significant training and accreditation is required to use particular lines of products or equipment. Beauty practitioners are much less likely to be regarded as professionals than medics and are classed as low-end service-providers; similar to retail workers they are

19.    Breast implants are now relatively routine with low complication rates (BAPPS, 2008). By contrast buttock lifts are far more risky (Widdows, forthcoming 2018). Moreover, some fairly frequently carried out surgeries, have startling high risks of serious complications. For example, as many as a quarter of those who have abdominoplasties (tummy tucks) require further surgery (Stewart et al., 2006).

20.    For example, fillers could be carried out at home, in a salon or in a medical setting.

21.    Non-surgical procedures promoted as being carried out by medical professionals may be done by dentists or GPs rather than experts in cosmetic work, likewise surgeons may not be specialised or experienced in the surgeries they are undertaking.

22.    Glen Jankowski makes a parallel argument with regard to fashion. He argues that while there has been lots of attention on the need of the fashion industry to address the rise in body image worries by using more diverse models and images, there has been very little attention to the injustices perpetrated on those who work in the 250 million sweatshops providing the clothes of the fashion industry (Jankowski, 2016).

regarded as selling a product and a service. They are not generally regarded as part of the professional classes. This is significant with regard to the power dynamics of the relationship between provider and recipient. The cosmetic surgeon is powerful, can suggest, explain, delimit and ultimately refuse to provide surgery. By contrast the beauty practitioner is expected to deliver what the client wishes, and while they can advise and suggest, ultimately it is the client's wishes which largely prevail. The power rests with the consumer, the providers' role is to meet the recipients' desires, making it hard for providers to control what they do; including the extent to which they work with and administer risky products.[23] There are of course exceptions and counter examples; celebrity hair-dressers are in high demand and they can charge significant amounts of money and dictate their clients' hairstyles. Likewise, surgeons can be in vulnerable positions (for example, surgeons from lower-income countries, contracted to cosmetic surgery companies who fly them in to perform a series of operations in a short period of time).[24] In such scenarios the surgeons do not meet with recipients in advance and therefore have little control over who they operate on or the operations they do. Moreover, even the most expert and professional surgeon might feel under pressure to deliver, or try to deliver, what the recipient asks for.[25] A frequent comment from cosmetic surgeons is that if they do not agree to operate the recipient will simply find another surgeon who will operate; and the implication is that this would be worse for the recipient. This said, for the most part, those who work as beauty professionals at the more routine, less medicalised, end of the spectrum have less power relative to the recipient, less control about what they deliver to a particular client, and less social and economic capital. Given the relatively low value placed on such beauty work, such workers often find themselves in highly competitive environments, vulnerable to being priced out of the market or replaced, especially given the relatively low barriers to entering the profession. As such they are low-paid and low-status workers vulnerable to exploitation. They are unable to ask for, or to provide for themselves, better pay and less harmful and risky working con-

23.    Debra Gimlin captures this difference between surgeons as powerful service providers, whose judgements are regarded as expert and who chose what they will and will not do, and hairdressers who seek to advise but are not treated as experts and ultimately do what the client wishes (Gimlin, 2002).

24.    The so called 'seagull surgeons' discussed in footnote 3.

25.    The narrative of seeking to provide what the recipient wants and yet being aware that some of those who seek cosmetic surgery and/or beauty interventions are looking for impossible transformations, is a prevalent discourse of both surgeons and beauticians. This arguably adds a further emotional pressure on providers.

ditions if they wish to keep their jobs or remain competitive. To illustrate, consider the harms which are regularly suffered by nail technicians.

Nail technicians suffer physical harm from working with toxic chemicals on a daily basis and from inhaling the dust produced by filing acrylic nails. The *New York Times* ran an expose of the conditions of nail technicians working in New York in the summer of 2015 (Nir, 2015a; Nir 2015b). These articles documented the experience of the women who work in nail bars and their experiences of illnesses caused by the chemicals and materials involved in providing acrylic nails and gel polish. The harms which the women reported either experiencing or being aware of included miscarriage, cancers, skin irritations and respiratory problems. They also reported that children are frequently born with health issues and learning difficulties. Of course such reports are not evidence-based. These conditions, which women either experienced or knew women who had experienced, were not verified, nor could their causes be directly equated to their working conditions (at least not without further research). However, "some of the chemicals in nail products are known to cause cancer; others have been linked to abnormal fetal development, miscarriages and other harm to reproductive health" (Nir 2005b), making this work *prima facie* risky. Such harms are not limited to nail technicians and there is evidence to suggest that hairdressers also run risks from frequent exposure to toxic chemicals (Takkouche, 2009).

The harm to providers of engagement in at least some beauty practices is direct and rarely mentioned. There are numerous reasons for this. Not least, the focus on cosmetic surgery in isolation from other beauty practices, means that often the providers are assumed to be relatively powerful and privileged surgeons; although less privileged medical professionals such as nurses are key deliverers of non-surgical practices. For obvious reasons surgeons are far less vulnerable to harm than hairdressers or nail technicians who come from poorer and more vulnerable demographics. Yet the harms to providers are direct and almost wholly missing from discussions about the consequences of the increasing value placed on appearance, the increased engagement in beauty practices, and the rise of minimal standards of beauty. If manicured nails become, as hair dye already is, part of the minimal requirement of beauty then the numbers of those being directly harmed will dramatically increase, as will the need for urgent action to address such direct harm.

*4.2 Indirect, specific harms to those who are 'abnormal'*

The second harm to others is indirect, but falls on a particular group of identifiable individuals; although exactly who falls into this group is open to discussion. The group in question is those who do not, and cannot conform to minimal beauty norms. This group is made up of individuals who fall significantly outside what is considered to be 'normal', or 'just good enough' when it comes to meeting beauty ideals. Those in this group obviously fail to meet beauty standards and in ways which will not be experienced by most individuals. For instance, they are not failing to measure up because they are old, overweight or hairy (all of which might be failing to meet appearance standards, but either in ways which can be addressed or in ways which are commonly, 'normally', experienced). Those in this group obviously, uncontroversially and perhaps permanently fall outside the normal range. Those in this group a can be termed 'abnormal'. I use the terms normal and abnormal reluctantly but knowingly and deliberately as if they are clear categories. My reason for doing so is to highlight the significance of the potential harm, and to show the serious nature of the risks which attach to the narrowing of normal in a context where appearance is increasingly valued.

Those who fall into the abnormal group are those who are disfigured at birth or by accident or have physical features which fall dramatically outside the normal range. What matters is that for this group there is no possibility of attaining minimal standards of normal and this is obviously and strikingly the case. To put it simply the individuals who fall in this group are clearly abnormal to the observer in a way which means that they fail to meet at least some aspect of minimal beauty norms; although they may, of course, meet other features of beauty norms.[26] My claim is that those who fall a long way outside the normal range will become more visible—and so more vulnerable—as beauty becomes more important and as minimal standards of beauty rise. As we 'fix' what can be fixed, the gap between what is normal and what is abnormal grows and the widening of this gap may harm those who fall into the abnormal category. The harms which such a group might suffer, taken broadly and indicatively, are: harms with regard to self-conception and identity (including lower self-esteem and increased feelings of shame and anxiety); harms of increased stigma and discrimination; and harms of exclusion. As appearance increasingly matters for presentation and communication in an increasingly visual and virtual world social exclusion is

---

26.    For example, smooth and unblemished skin may be attainable for an individual who cannot attain beauty norms of thinness and height and conversely someone with disfigured skin (for example as a result of acid attack) may attain thinness, and so on.

a real risk for those who cannot meet the appearance norms of the digital world or who are uncomfortable with communication which is primarily image- rather than text-based. Indeed, as participation in, and being 'liked' on virtual image-based platforms grow, meeting appearance thresholds might become an effective precondition of social interaction.

Both of these claims—first that this group will be more different and visible and second that they will be harmed as a result of increased difference and visibility—are open to critique, and in part empirical critique. In this paper it is not possible to definitively claim that this group are (or will be) harmed by increasing minimal standards of beauty, but it is possible to argue for their increased visibility and to suggest that this makes harm a real and reasonable possibility. To be clear, the widening of the gap between normal and abnormal does not necessarily result in harm, but it does make it possible for harm to occur in comparison to a scenario where the gap between what is normal and abnormal is narrower or indecipherable or where measuring up to normal appearance ideas is not regarded as socially or culturally valuable. I will consider first the narrowing of normal and extending the distance between normal and abnormal and second why this might result in harm for this group.

First, the narrowing of normal. That more is required to attain minimal standards of beauty, to be 'good enough' or to 'be normal' is an argument I make in depth and at some length elsewhere. In brief I argue that increasingly more is required of more of us to attain minimal standards of beauty and that this is happening incrementally and in some instances what is required to meet minimal standards has risen dramatically over a short period of time and with little critique.[27] Examples at the routine end of the spectrum include body-hair removal, where in a generation visible body hair has gone from acceptable—and even sexy—to unacceptable and shameful. Another practice which was once occasional or optional but is now required to meet minimal standards of beauty (particularly for women, but increasingly for men) is hair dye; most women over a certain age, dye their hair. Further, the increase in nail bars suggests that manicures might be the next practice to tip into the required category.[28] In some contexts and among some demographics minimal standards go well beyond these. Already, in some circles Botox and fillers fall into the required category and it is the untreated and aging face which is abnormal (Kay, 2015). Moreover, while

27.    I argue that this is globally the case and that this matters for the normalising claim as there are less competitor norms (Widdows, forthcoming 2018).

28.    I make this argument in detail in chapter 4, "Routine maintenance, Treats and Extremes", of *Perfect Me* (Widdows, forthcoming 2018).

cosmetic surgery is still exceptional in most groups and contexts it is increasingly desired and normalised.

The gradual escalation of minimal standards, coupled with increased pressure to engage beyond what is minimal and towards what is maximal, results in a narrowing of the normal range. As those who can engage to attain ever more demanding minimal standards the group I am concerned with become increasingly differentiated from the norm; more visible and more different. In addition, those in this group may also become rarer, again making them more visible. As those things which can be 'fixed' are fixed in order to attain as near an approximation to normal as possible those who cannot be 'fixed' stand out more. There are numerous examples of 'fixing' features routinely, often for supposed health reasons, even though health functioning is not affected. To illustrate, sticky-out ears are routinely 'pinned back' for appearance reasons even though hearing and other health functioning is not affected, likewise birthmarks and other disfigurements are removed. Similarly, if affordable, teeth are routinely straightened, and while some of this work is for functioning reasons, for instance, to reduce overbite or/and to reduce plaque build-up, much teeth straightening is for appearance reasons.

In addition to fixing abnormalities, it is also increasingly the case that appearance is regarded as a reason for pre-natal selection (whether by Pre-Implantation Genetic Diagnosis or pre-natal testing followed by termination). For instance, while the numbers of abortions to avoid a child with a cleft lip or palate are small (McHale and Jones, 2012), and smaller than the hype would have us believe, such terminations do occur and presumably the reasons for such choices are appearance based. While the functioning of a cleft palate or lip can usually be corrected by surgery often the tell-tale signs remain. Such features are disfigurements and as appearance matters more they are becoming regarded as significant; and significant enough to justify selection against. This suggests that appearance is a factor in decisions about selection and perhaps to the extent where a flawed appearance, irrespective of functioning, is regarded as a significant impartment. Selection on appearance grounds is likely to extend, for example, it has been claimed that were the technology available 11 percent of couples would abort a foetus predisposed to obesity (Rhode, 2010, 26). I do not wish to enter discussions about whether or not those who are selected against are harmed, but I do wish to argue that increased selection makes those in the abnormal category rarer, more visible and as a result makes them more open to harm. In addi-

tion such judgements contribute to the belief that appearance is highly valuable and accordingly that failure to succeed in attaining beauty goods is significant.

Given these trends it is not unreasonable to think that those who fall outside the normal range will indeed become rarer and as a consequence more visible. Yet being visible or rare, even to the point of abnormality, is not necessarily a harm, which brings me to the second argument. To argue that such individuals are, or will be, more open to being harmed requires an additional, and more difficult, argument. It requires empirical evidence which is simply not available with regard to current harm. However, while the evidence is not yet sufficient to prove assertions of current harm, that harms are possible, perhaps likely, is sufficient reason for policy makers to be alert to the possibility and to seek to establish whether such harm is occurring.

Looking forward it is possible that increased rarity and visibility of this group might result in increased vulnerability and harm, but the opposite may also prove to be true. On the one hand it is the case, at least in very many places, that those who are very visibly disfigured or abnormal are in far better situations than they were in previous generations; following disability rights activism and regulation of the last half-century (Shakespeare, 2006). It is not just the case that there is less discrimination permitted in certain sectors, for example, in terms of employment opportunities, but also that successful activism has resulted in cultural change with regard to those who are visibly disfigured; in the terminology of this paper abnormal. Certain types of harm—most obviously discrimination, but also stigma and shaming—are not just illegal but unacceptable. This is clearly evident in terms of acceptable language and behaviour. This tolerance may mean that even if there is increased discrimination on appearance grounds this group may not suffer from it; because discrimination of the obviously physically disabled or disfigured is unacceptable. Indeed it might be the case that those who fall a long way outside the normal range might prove less vulnerable to harm than those who fall nearer to it. The impossibility of attaining normal might protect those in the abnormal group from pressures to attain appearance norms all together. On the other hand it is reasonable to think that as appearance matters more, the suffering of those who do not make the appearance grade will increase. If meeting beauty ideals becomes an increasingly valued aspect of personal identity and a key factor is self-esteem then not being able to meet these criteria will be increasingly costly with regard to identity and self-conception, as well as carry costs with regard to being valued in the public sphere and able to access other goods, such as employment. Exactly how such harms might accrue and be constituted is not

clear and the extent to which harms will in fact manifest requires further evidence. However, that this group are more visible is clear and, given this, it is not unreasonable to think this makes them open to suffering a number of possible harms. This deserves both recognition and attention.

### 4.3 Indirect, general harms to all

The third harm is the indirect harm to all of rising standards of beauty and increased engagement with beauty practices. As more is required to meet minimum standards more of us fall short and fail. Arguably this leads to an increasingly toxic environment which constitutes a general harm to all. In the last section I argued that rising minimal standards of beauty make those who fall a long way from the normal range more visible and different and potentially open to harm. In this section I argue that rising minimal standards not only harm those who are abnormal, who can never meet the increasing demands of beauty, but also those who can. As more engage so non-engagement stands out and becomes unusual and eventually abnormal; in the words of Susan Bordo "the ordinary body becomes the defective body" (Bordo, 1997). For example, it is the hairy body and non-coloured hair which are now abnormal. Beauty requirements are enforced by social norms and expectations, rather than by coercion, but nonetheless the list of beauty practices which are regarded as required is increasingly extensive and demanding. Rising minimal standards of beauty eventually fall on all who are able to conform, as not conforming becomes unacceptable; a failure to minimally groom connotes lack of respect for the self or others and signals illness or distress. Over time the choices of some to engage eventually mean that all have less choice not to engage. Thus the choices of some gradually reduce the choices of all.

This argument has parallels with the feminist argument discussed in 2.2; and in a similar manner seeks to suggest that social norms harm those who do not engage and potentially all. However, while there are some similarities—most importantly that the choices of individuals should not be considered as isolated or discrete—my claim is substantively different. The feminist arguments focus on the harms which arise from gender injustice; whether to individual women or to women as a group. Key to these claims is the asymmetrical and hierarchical relations between men and women and the difference between men and women's engagement in beauty practices. These arguments are not primarily concerned that beauty standards are increasingly demand-

ing, although this is part of the critique, but that they are demanded of women and not of men. According to such theories the purpose of beauty practices is to mark women out, to trivialise and sexualise and to embed unequal and inferior status. My target is not gender disparity or inequality, but the harms which accompany the increasing valuing of appearance in a visual and virtual culture and the harms which follow from increasing engagement in beauty practices and the accompanying rise in minimal standards of beauty. Such rising engagement is required of all, irrespective of gender.

There are of course valid arguments about the gendered harms of beauty, but the harms which follow from rising minimal standards are not necessarily gendered, nor do they derive from asymmetry between men and women. Elsewhere I have argued in detail that the asymmetry, or inequality to use the more familiar language of moral and political philosophy, between genders with regard to body work is breaking down. Men, particularly young men, increasingly do body work and engage in beauty practices to attain normal or minimal standards. They also increasingly suffer, as women long have, from body image anxiety, and experience pressure to attain body ideals which require significant time and effort and often chemical and/or surgical intervention.[29] Accordingly, for all, irrespective of gender, what is required to attain minimal standards of beauty, to be just good enough, to be normal, is growing. Beauty requirements are more demanding, and there is more pressure to attain minimal standards, and more pressure to engage beyond the minimum for both men and women. Moreover, and crucially for this paper, more harms attach if one fails to attain such minimum standards, or believes that one has failed. The types of harms which attach to failing in the appearance stakes, or simply to over-focusing on appearance, are myriad. They include harms to individuals as anxiety about body image and appearance increases and increasingly beauty failure is regarded as failure more generally. In addition there are shared and communal harms as time, effort and money is devoted to beauty goods in preference to other goods. This harm is a standard justice harm of opportunity cost: what could we do if we did not do this? Given the size of the beauty industry, taking into account the use of global resources, as well

---

29.    To clarify, I argue that the inequality between men and women with regard to beauty is breaking down in four key ways: first, men increasingly value (male) appearance; second men increasingly worrying about body image; third men increasingly engage in body work; and fourth male beauty ideals, like female beauty ideals, are increasingly demanding and require intervention (Widdows, forthcoming 2018). However, significant gender differences – and harms from such gender differences – remain.

as expertise (for instance, R&D devoted to beauty which could be devoted to health), the collective opportunity cost is extensive.

However, while it is the case that the harms from an increasingly toxic environment are potentially devastating to individuals and communities it is not the case that these harms track directly to engagement of some in beauty practices. My claim is not that those who engage in beauty practices are wholly responsible for the rising minimal standards of beauty, nor am I arguing that those who engage should be blamed or prohibited from engagement.[30] Rather, I argue that it is reasonable to infer that rising engagement in beauty practices is one factor which contributes to a culture in which beauty standards will continue to rise, and in which normal will be harder to attain, and in which beauty will be increasingly valued. The harms of this are not, or not only, to those who engage, but to others who currently do not engage or who only engage to reach the most minimal standards of acceptable grooming.

Tracking the causes of social and communal harm is notoriously difficult and verges on the impossible. Undoubtedly increased engagement by some is only part of the picture. What is driving the increasing demands of beauty is complex and there are myriad reasons for increasing engagement, rising minimal standards and the increased value placed on appearance. Reasons include the rise of a virtual and therefore visual culture, technological advances which make new interventions possible, the increased democratisation of beauty as practices become affordable and accessible, and the dominance of consumer culture which values work on the self and prioritises body-projects as cites of self-expression and realisation (Jones, 2008; Gimlin, 2002; Lazar, 2011; Tincknell, 2011). While all agree that there is an exponential rise in anxiety about body image and increased valuing of appearance attributing causes is complex. Yet for whatever reason body image is cited as the third largest and most harmful challenge facing young people in the UK (after lack of employment opportunities and failing to succeed within the education system) (YMCA, 2016). The literature on which factors are most important in feeding this rise in body image anxiety is large, contested and indeterminate.[31] But, while it is not possible to track the extent to which each factor contributes, it is the case that appearance increas-

---

30.    On the contrary, I argue that rejecting beauty practices is divisive, counter-productive and is a failed response to the growing demands of beauty (Widdows, forthcoming 2018).

31.    For example, some focus on the media, arguing that the link between idealised thinness in the media and body images issues is now demonstrated in many studies (Ghaznavi and Taylor, 2015). While others deny media influence and argue that it is the influence of family and friends which is primary (Stice et al., 2001).

ingly matters, particularly to the young, and, as tracking changes in beauty practices shows what is required to attain minimal standards is more demanding than previously. Undoubtedly then, "the high prevalence of negative body image is a significant public health concern due to its negative physical and psychological health outcomes" (Diedrichs et al., 2011).[32] Overall the rise in attention to beauty, engagement in beauty practices and rising minimal standards results in significant communal or general harm which falls on all. We are harmed by the creation of a toxic environment in which appearance is dominant, cosmetic surgery is normalised, and beauty is key to personal identity and denotes individual success and failure.

To argue that an increasingly toxic environment is emerging and that this harms others (all others), is not to suggest that all are harmed in the same way or to the same extent and some individuals may be virtually unscathed. Individuals will respond differently to different types of stimuli and negotiate and critique such pressures differently and some will be unaffected. Yet the impossibility of determining causal links or of tracking direct impact should not mean we do not pay attention to patterns and the potential harmful consequences of changing social norms and contexts. Communal harms matter not in an abstract way but precisely because they impact on all by shaping and limiting what is possible for individuals to be and do.

## 5. CONCLUSION

In this paper I have argued for a shift in framework in order to recognise currently neglected harms of beauty. I have argued that current discourses, especially those of policy-makers, ethicists and lawyers, have largely ignored harms-to-others. They have focused almost exclusively on the harms to those who engage beyond minimal standards of beauty. I argued that this is strange given the liberal assumption that harms-to-others should be addressed in preference to addressing harms to the self; as interference in self-harm is deemed wrongly paternalistic. I then turned to the harms-to-others which are currently omitted from the discourse surrounding beauty practices and I focused on three harms-to-others to illustrate the range of harms and the extent of such harms. The extent to which these harms-to-others are attributable and can be tracked differs. The first harm, to providers of beauty practices, is a direct

---

32.    Negative effects which have been suggested result from body image anxiety include (amongst others) lower self-esteem, disordered eating behaviours and eating disorders, impaired social and occupation functioning and well as poorer day-to-day interactions and increased problems with sexual functioning (Cash and Smolak, 2011).

harm, the second two are indirect and the extent to which engagement is causal and the extent of the harm is more difficult to evidence, but nonetheless such harms should be taken into account when considering the costs and harms of engagement in beauty practices.

The significance of the harm, rather than type of harm, should determine the extent to which policy-makers seek to intervene. The current prevalence of body image anxiety experienced, particularly but not only, by young people and its negative effects (increased anxiety, low self-esteem, reduced physical activity, and lower social and educational involvement) is of epidemic proportions.[33] If these effects could be tracked to a physical cause, for instance the taking of a recreational drug, or as a side-effect of a pollutant, then such causes would immediately be targeted. How harms are recognised and the construction of harms fundamentally shapes policy responses. The classic example is the shift in the construction of the harms of smoking; once regarded a matter of individual freedom and choice and now regarded as a public health issue. A currently contested example which is increasingly regarded as a public health issue is obesity; although there is still significant debate about the extent of justified State and policy intervention. If the same reframing happened with beauty, and the harm of a toxic environment which has created an epidemic of body image anxiety, was reframed as a public health issue, rather than a matter of individual choice, policy would be transformed. No longer would this be regarded as something individuals should be left to choose to do or not do, but harm reduction would be introduced. A host of interventions are possible, including prohibiting some practices (at least for some groups), regulating advertising, providing education, making social media accountable and so on. How effective such interventions are likely to be requires further research, but what intervention will work matters less than recognising that we should intervene. Recasting beauty harms as public health concerns provides a reason, even a duty, to intervene. It reveals the harms which focusing only on engaging individuals obscures. Admittedly targeting social causes and addressing communal harms is far more difficult than regulating individual choices, however difficulty in addressing harms should not prevent harms from being recog-

---

33.    There is numerous evidence to support this claim, including the YMCA report (2016) and the annual Girlguiding survey (2016). A few statistics from the Girlguding survey are indicative: 47 percent of girls aged 11-21 say the way they look holds them back; 40 percent of girls between 7 and 10 think feel they should lose weight sometimes or most of the time and this rises to 80 percent of girls between 17 and 21; 53 percent of girls.

nised. Policy should focus on addressing the most significant harms not those which are easiest to address. Focusing on engaging others results in a skewed picture which makes significant and potentially devastating harms of beauty invisible.

REFERENCES

ASAPS (American Society for Aesthetic Plastic Surgery) (2016). "Cosmetic Surgery National Data Bank Statistics." Available from http://www.surgery.org/sites/default/files/ASAPS-Stats2016.pdf [Accessed 28 April 2017].

BAAPS (British Associaiton of Aesthectic and Plastic Surgeons) (2008). "Surgeons reveal UK's largest-ever Breast Augmentation survey." Available from http://baaps.org.uk/about-us/press-releases/404-surgeons-reveal-uks-largest-ever-breast-augmentation-surve. [Accessed 9 April 2017].

Bartky, Sandra Lee. (1990). *Femininity and Domination: Studies in the Phenomenology of Oppression.* New York and London: Routledge.

Blum,Virgina L. (2003). *Flesh Wounds: The Culture of Cosmetic Surgery.* Berkley, California and London: University of California Press.

Bordo, Susan. (1997). *Twilight Zones: The Hidden Life of Cultural Images from Plato to O.J..* Berkeley, Los Angeles, London: University of California Press.

Cash, T. F., and Smolak, L. (Eds.). (2011). *Body image: A handbook of science, practice, and prevention.* New York: Guilford Press.

Chambers, Clare. (2008). *Sex, Culture and Justice: The Limits of Choice.* University Park, PA: The Pennsylvania State University Press.

Corbett, William. R. (2007). "The ugly truth about appearance discrimination and the beauty of our employment discrimination law." *Duke Journal of Gender Law and Policy,* 14, 153-178.

Davis, Kathy (1995) *Reshaping the Female Body: The Dilemma of Cosmetic Surgery*. New York and London: Routledge

Diedrichs, Phillippa C. Christina Lee and Marguerite Kelly. (2011). "Seeing the beauty in everyday people: A qualitative study of young Australians' opinions on body image, the mass media and models." *Body Image* 8(3), 259-266.

Dworkin, Andrea. (1974). *Women Hating.* New York: E. P. Dutton.

Eagly, Alice H., Richard D Ashmore, Mona G Makhijani and Laura C Longo. (1991). "What is beautiful is good, but.....: a meta-analytic review of research on the physical attractiveness stereotype." *Psychological Bulletin*. 110(1), 109-128.

Edmonds, Alexander. (2007). "The poor have the right to be beautiful': cosmetic surgery in neo-liberal Brazil." *Journal of the Royal Anthropological Institute*, 13(2), 363-381.

———(2010). *Pretty Modern: Beauty, Sex, and Plastic Surgery in Brazil.* Durham and London: Duke University Press.

Etcoff, Nancy. (1999). *Survival of the Prettiest.* New York: Anchor Books.

General Medical Council (GMC). (2016). *Guidance for Doctors who offer Cosmetic interventions*. Available from https://www.gmc-uk.org/Guidance_for_doctors_who_offer_cosmetic_interventions_210316.pdf_65254111.pdf [Accessed 9 April 2017].

Ghaznavi , Jannath, and Laramie D. Taylor. (2015). "Bones, Body Parts, and Sex Appeal: An Analysis of #thinspiration Images on Popular Social Media." *Body Image* 14, 54-61.

Girlguiding (2016). *The Girl's Attitude Survey*. Available from https://www.girlguiding.org.uk/social-action-advocacy-and-campaigns/research/girls-attitudes-survey/ . [Accessed 10 April 2017].

Gimlin, Debra. (2002). *Body work: Beauty and self-image in American culture.* Berkeley, Los Angeles and London: University of California Press.

———(2012). *Cosmetic surgery narratives: A cross-cultural analysis of women's accounts.* Basingstoke: Palgrave Macmillan.

Griffiths, Danielle, and Mullock, Alex. (2017). "Cosmetic Surgery: Regulatory Challenges in a Global Beauty Market." *Health Care Analysis* (online first).

Hamermesh, Daniel S. (2011). *Beauty Pays: Why Attractive People are More Successful*. Princeton and Oxford: Princeton University Press.

Jankowski, G. S. (2016). "Who stops the sweatshops? Our neglect of the injustice of maldistribution." *Social and Personality Psychology Compass*, 10(11), 580-591.

Jeffreys, Sheila. (2005). *Beauty and Misogyny: Harmful Cultural Practices in the West*. London and New York: Routledge.

Jones, Meredith. (2008). *Skintight: An Anatomy of Cosmetic Surgery*. Oxford and New York: Berg.

Kay, Karen (2015). "Is Cosmetic Surgery the new acceptable face of womanhood." *The Guardian*. Available from http://www.theguardian.com/lifeandstyle/2015/jun/28/cosmetic-surgery-normal-acceptable-face-womanhood. [Accessed 9 April 2017].

Keogh, B. (2012). *Poly Implant Prothese (PIP) breast implants: Final report of the expert group*. Department of Health. Available from

https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/214975/dh_134657.pdf [Accessed 9 April 2017].

———(2013). *Review of the regulation of cosmetic interventions*. (Keogh Review) London: Department of Health. Available from https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/192028/Review_of_the_Regulation_of_Cosmetic_Interventions.pdf [Accessed 9 April 2017].

Laimann, Jessica. (2015). "Should we prohibit breast implants? Collective moral obligations in the context of harmful and discriminatory social norms." *Journal of Practical Ethics*, 3(2), 37-60.

Lazar, Michelle M. (2011). "The Right to be Beautiful: Postfeminist Identity and Consumer Beauty Advertising." In *New Femininities: Postfeminism, Neoliberalism and Subjectivity*, edited by Rosalind Gill and Christina Scharff , 37-51. Basingstoke: Palgrave Macmillan.

McHale, Jean V., and June Jones. (2012). "Privacy, confidentiality and abortion statistics: a question of public interest?" *Journal of Medical Ethics* 38(1), 31-34.

Manson, Neil and O'Neill, Onora. (2007). *Rethinking Informed Consent in Bioethics*. Cambridge: Cambridge University Press

Nir, Sarah Maslin (2015a). "The Price of Nice Nails." *The New York Times*. Available from http://www.nytimes.com/2015/05/10/nyregion/at-nail-salons-in-nyc-manicurists-are-underpaid-and-unprotected.html [Accessed 3 August 2016].

———(2015b). "Perfect Nails, Poisoned Workers." *The New York Times*. Available from http://www.nytimes.com/2015/05/11/nyregion/nail-salon-workers-in-nyc-face-hazardous-chemicals.html [Accessed 3 August 2016].

Nuffield Council on Bioethics, (2017) "*The ethics of cosmetic procedures*." Available from http://nuffieldbioethics.org/wp-content/uploads/Cosmetic-procedures-full-report.pdf [Accessed 28 October 2017].

Rhode, Deborah L. (2010). *The Beauty Bias: The Injustice of Appearance in Law and Life*. Oxford and New York: Oxford University Press.

Royal College of Surgeons (RCS), Professional Clinical Standards. (2016). "Professional Standards for Cosmetic Surgery." Available from https://www.rcseng.ac.uk/library-and-publications/college-publications/docs/professional-cosmetic-surgery/. [Accessed 9 April 2017].

Shakespeare, Tom. (2006). *Disability Rights and Wrongs*. London and New York: Routledge.

Stewart, K. J., Stewart, D. A., Coghlan, B., Harrison, D. H., Jones, B. M., and Waterhouse, N. (2006). "Complications of 278 consecutive abdominoplasties." *Journal of Plastic, Reconstructive and Aesthetic Surgery*, 59(11), 1152-1155.

Stice, Eric, Diane Spangler and W. Stewart Agras. (2001). "Exposure to Media-Portrayed Thin-Ideal Images Adversely affects Vulnerable Girls: A Longitudinal Experiment." *Journal of Social and Clinical Psychology* 20(3), 270-288.

Swami, Viren, and Adrian Furnham. (2008). *The Psychology of Physical Attraction*. London and New York: Routledge.

Takkouche, Bahi, Carlos Regueira-Méndez, and Agustín Montes-Martínez. (2009). "Risk of cancer among hairdressers and related workers: a meta-analysis." *International Journal of Epidemiology* 38 (6), 1512-1531.

Tincknell, Estella. (2011). "Scourging the Abject Body: Ten Years Younger and Fragmented Femininity Under Neoliberalism." In *New Femininities: Postfeminism, Neoliberalism and Subjectivity*, edited by Rosalind Gill and Christina Scharff, 83-95. Basingstoke: Palgrave Macmillan.

Vartanian, L. R. (2009). "When the body defines the self: Self-concept clarity, internalization, and body image." *Journal of Social and Clinical Psychology*, 28(1), 94-126.

Widdows, Heather. (2017) "Exploitation and the Global Demands of Beauty." In *Exploitation: From Practice to Theory*, edited by Monique Deveaux and Vida Panitch London and New York: Rowman & Littlefield International, 179-193

———(forthcoming 2018) *Perfect Me!* Princeton, Princeton University Press.

World Health Organisation (WHO), 2011, "Mercury in Skin Lightening Products." Available from http://www.who.int/ipcs/assessment/public_health/mercury_flyer.pdf [Accessed 20 August 2017].

YMCA (2016) "The Challenge of being Young in Modern Britain." Available from http://www.ymca.co.uk/campaigns/world-of-good. [Accessed 9 April 2017].

# Why be Moral in a Virtual World?

JOHN MCMILLAN

*University of Otago*

MIKE KING

*University of Otago*

## ABSTRACT

This article considers two related and fundamental issues about morality in a virtual world. The first is whether the anonymity that is a feature of virtual worlds can shed light upon whether people are moral when they can act with impunity. The second issue is whether there are any moral obligations in a virtual world and if so what they might be.

Our reasons for being good are fundamental to understanding what it is that makes us moral or indeed whether any of us truly are moral. Plato grapples with this problem in book two of *The Republic* where Socrates is challenged by his brothers Adeimantus and Glaucon. They argue that people are moral only because of the costs to them of being immoral; the external constraints of morality.

Glaucon asks us to imagine a magical ring that enables its wearers to become invisible and capable of acting anonymously. The ring is in some respects analogous to the possibilities created by online virtual worlds such as Second Life, so the dialogue is our entry point into considering morality within these worlds. These worlds are three dimensional user created environments where people control avatars and live virtual lives. As well as being an important social phenomenon, virtual worlds and what people chose to do in them can shed light on what people will do when they can act without fear of normal sanction.

This paper begins by explaining the traditional challenge to morality posed by Plato, relating this to conduct in virtual worlds. Then the paper will consider the following skeptical objection. A precondition of all moral requirements is the ability to act. There are no moral requirements in virtual worlds because they are virtual and it is impossible to act in a virtual world. Because avatars do not have real bodies and the

persons controlling avatars are not truly embodied, it is impossible for people to truly act in a virtual world. We will show that it is possible to perform some actions and suggest a number of moral requirements that might plausibly be thought to result. Because avatars cannot feel physical pain or pleasure these moral requirements are interestingly different from those of real life. Hume's arguments for why we should be moral apply to virtual worlds and we conclude by considering how this explains why morality exists in these environments.

<p style="text-align:center">———— •◦• ————</p>

## INTRODUCTION

Our reasons for being good are fundamental to understanding what it is that makes us moral or indeed whether any of us truly are moral. Plato grapples with this problem in book two of *The Republic* where Socrates is challenged by his brothers Adeimantus and Glaucon (Plato 1993). They argue that people are moral only because of the costs to them of being immoral; the external constraints of morality.

Glaucon asks us to imagine a magical ring that enables its wearers to become invisible at will, and capable of acting anonymously. He relates a fiction in which a shepherd named Gyges discovers this ring and uses its powers to seduce the queen of the kingdom, kill her husband, and take control of the kingdom. Glaucon claims that if there were two such 'rings of Gyges' and one was worn by a previously moral person and the other worn by a previously immoral person that the moral person would end up committing immoral actions too. So, the central point of this thought experiment is to claim that people only do the right thing because of the potential rewards of identifiable right action and the potential punishments of wrong action, which we refer to as the sanctions of morality. The internal constraints of morality or moral reasons are weak or nonexistent, causally ineffective.

The ring of Gyges enables its wearers to act without fear of detection and punishment. It doesn't make them omnipotent or omniscient. However, it is in some relevant respects analogous to the possibilities created by online virtual worlds such as Second Life. These worlds are three dimensional partially user-created environments where people who are members of the social network control avatars that 'live' virtual lives. These avatars and the 'lives' they lead need bear no relation at all to

the person controlling them, or their life outside of the virtual world. Avatars can perform a very wide range of actions, interact with others in the virtual word, attend lectures or performances, and engage in many other activities.

Launched in 2003 by Linden Labs, Second Life is one of the earliest and arguably most successful online virtual worlds, with an estimated 1 million regular users in recent years (Levy 2014). Linden Labs has recently announced Project Sansar, which claims to give a higher level of freedom to users to create their own highly detailed virtual content, and to incorporate virtual reality headset technology to create a more realistic experience for users (Linden Labs 2016a). As well as virtual worlds being a very important social phenomenon, the way people use their avatars within these worlds may shed light on what people might do when fear of sanction is diminished.

This paper begins by explaining the traditional challenge to morality as posed by Plato. Then it considers whether the anonymity and avoidance of external sanction possible in virtual worlds is a useful test case for the challenge to Socrates. We argue that although virtual worlds do exhibit the problems of reduced prudential reason to be moral, virtual worlds raise more acutely the question of whether there is any moral reason to act beyond mere prudence. Then the paper will consider the following skeptical objection. As Kant observed, a precondition of all moral requirements is the ability to act. The acts of avatars occurring within virtual worlds are not, and cannot be, acts in the sense intended by Kant. Since it is impossible to act in a morally relevant way in a virtual world, there can be no moral requirements constraining the actions of avatars in virtual worlds. We counter this objection by arguing that it is possible for avatars to act in ways that are relevant for morality, and suggest a number of moral requirements that might plausibly be thought to result. However, since avatars are different from physically embodied people in morally relevant ways these moral requirements are interestingly different from those of real life. While immoral actions such as rape lack some of their physical consequences in a virtual world, their psychological impact and what they express about the attitudes of those who perform them, are good reasons for viewing them as immoral.

Generating an account of the moral requirements of a virtual world is no less difficult than creating an account of those in the actual world. However, in both cases an important place to start is by considering the nature of action. Taxonomizing the actions that are possible in a virtual world is a significant project in its own right and this paper will confine itself to showing how three actions that would be immoral if committed against people in the actual world (murder, rape and slavery) are not the

same kind of action in a virtual world. Then, it will defend two instances of action (promising and asserting) that are similar in actual and virtual worlds. In each case, the relevance of the nature of the act for moral requirements relating to that act will be considered.

We will suggest that, contrary to what we might expect given the nature of virtual worlds, morality can exist and flourish within them. If this is so, then the question 'why be moral in a virtual world?' can be subsumed within the more general question, 'why be moral?' If there is reason to be moral, then this reason will retain its normative strength within virtual worlds. The final part of the paper revisits David Hume's discussion of the 'sensible knave' and argues that those who act morally in a virtual world experience what Hume calls the 'invaluable enjoyment of a character' and those who do not have abandoned this for the sake of a few worthless virtual gewgaws. The Humean observation also explains why it is that morality appears to flourish, albeit in a different form, in virtual worlds.

## THE RING OF GYGES

The challenge to Socrates is skepticism about whether those who appear to act on moral reasons are genuinely acting on moral reasons. Socrates is asked to consider how a shepherd would act if he found a ring that enabled him to become invisible. The story goes that upon discovering the power of the ring the shepherd seduced the Queen and murdered the King so that he might take the throne.

It might be objected that the immorality of the shepherd is due to moral weakness on his part and someone moral would have acted otherwise. So as to rule out this possibility, Glaucon introduces the following extension of his thought experiment.

> *Suppose there were two such rings, then—one worn by our moral person, the other by the immoral person. There is no one, on this view, who is iron-willed enough to maintain his morality and find the strength of purpose to keep his hands off what doesn't belong to him, when he is able to take whatever he wants from the market-stall without fear of being discovered, to enter houses and sleep with whomever he chooses, to kill and to release from prison anyone he wants, and generally to act like a god among men. His behaviours would be identical to that of the other person: both of them would be heading in the same direction (Plato 1993, pp 47-48).*

Glaucon then claims that people never freely choose to act morally; if given the option of acting immorally and doing what was in our interests, without fear of detection, we would act immorally because acting morally is a burden and not prudentially valuable.

One reason that this is such an elegant thought experiment is because if Glaucon's predictions are correct it can account for the fact that many people appear to act morally. The appeal to external sanctions is consistent with the common everyday observation that those who have less to lose are, all other things being equal, more likely to commit a crime. An alcoholic who lives on the street has different reasons for not stealing a bottle of bourbon than the high school teacher who finds herself out of cash and in need of a drink. Glaucon's claim suggests that those who follow the rules of morality do so because they judge the costs of noncompliance to be too great and this seems reasonable given that those costs can differ depending upon the person.

The ring of Gyges enables its wearers to act anonymously and without fear of detection and punishment. It doesn't make its wearers godlike in other ways so it's magic is not directly comparable with other corrupting powers in fiction such as the One Ring in *The Lord of the Rings*. Glaucon claims that mere anonymity is sufficient for revealing the true nature of morality and that all apparent moral reasons are reducible to self-interest.

While this is a thought experiment it is not simply an ethical intuition pump. In effect, it is an empirical claim about our psychology and what we would do if placed in a situation where we could act without fear of punishment or criticism. Whether Glaucon is correct depends upon whether it really is the case that this is what we would do, if we could act free from the threat of punishment or criticism.

Wartime atrocities, especially those that have occurred when soldiers believed they were following orders or immune from punishment, show that many apparently normal human beings are capable of appalling actions. While alarming, immoral behavior of that kind doesn't give direct support for Glaucon's prediction because they are only committed by a select group of persons, but also because wartime atrocities occur under a set of unusual conditions. Jonathan Glover describes the process of 'moral distancing' whereby military actors become increasingly alienated from the moral quality of their actions (Glover 2000). This means that those who commit atrocities during wartime may not do so simply because the external sanctions of morality have been removed but also because their responsiveness to moral reasons

has been eroded. It is also possible that those who commit wartime atrocities do so because they don't see what they are doing as wrong, or think that they have been ordered to do so, as was a feature of some of the abuse at Abu Graibh (Brown 2005).

Milgram's obedience studies are better evidence because they show that many people during peacetime would, if ordered to do so by an authority figure, cause grave harm to another person (Milgram 1963). It's unclear whether the experimental subjects that acted immorally did so because they believed that the presence of an authority figure meant they would not be punished. Nonetheless, it seems reasonable to suppose that many of them must have thought this (Gillett and Pigden 1996). Some subjects believed that they gave another person a lethal shock. If the subjects thought that they could be charged with murder it would be irrational, as well as immoral, to give this shock even when being asked to do so by a man in a white coat. Milgram varied the study by removing the 'instructor' from the room so that they gave the research subject instructions via the phone, and thereby could exercise more freedom over the extent to which 'learners' were shocked.

While the Milgram experiment shows that an alarming number of people will act immorally if ordered to, it doesn't show that all persons will act immorally of their own accord, as Glaucon claims. The experiment was designed to test the extent to which an authority figure could influence behavior so isn't an instance where persons can act in whichever manner they chose without fear for the consequences of acting immorally.

Social psychologists interested in the proclivity of men to rape have studied the effect of beliefs about punishment on the likelihood of rape. Some studies have suggested that many men would rape if they believed there was no chance of them being caught (Malamuth 1981). Alarming as these findings are, they are complicated by the possibility that there could be a difference between what men say they would do, and what they would do under these conditions.

While these and other examples suggest that human beings are capable of immoral actions when the usual external sanctions of morality are altered, they are specific to particular actions and are complicated by the context that these people acted in.

Nogami and Takai found that in an experimental game setting, players that were anonymous (and therefore non-accountable and non-identifiable) broke rules to gain monetary reward (Nogami and Takai 2008). In the same game, rule-breaking to gain monetary reward did not occur in players who were only non-identifiable, only non-

accountable, or non-anonymous (both identifiable and accountable). This suggests that anonymity and the removal of external constraint is a critical factor in determining whether people will be immoral.

Virtual worlds enable people to develop new appearances and identities, in effect they can present themselves as a radically different kind of person. While it is easy for an avatar to reveal their actual world identity, the majority take advantage of the chance to be anonymous. This means that virtual worlds can create one of the preconditions of Glaucon's challenge: anonymity is analogous to a ring of invisibility, and what this provides is non-identifiability and non-accountability for actors in virtual worlds. This provides a reason to suspect that moral conduct in virtual worlds (and other settings in which non-identifiability and non-accountability is permitted) may be worse than that in real life. The evidence we have presented here supports this hypothesis.

However, the discussion of wartime atrocities by Glover also suggests strongly that wrongful acts by moral agents may be rationalized through undermining of the moral reasons that count against these acts. In virtual worlds this may be more acute, since the moral qualities of acts in virtual worlds is genuinely uncertain. Not only is the prudential reason provided by sanction weakened, but it is an open question as to whether any other moral value obtains in the virtual world, given its difference from the real world.

## VIRTUAL WORLDS: SECOND LIFE

There are a number of online virtual worlds. They differ in their size, number of residents, language and theme. The largest, most global and most relevant for our purposes here is the virtual world owned and supported by the San Francisco based company Linden Labs called Second Life. This virtual world is importantly different from predecessors such as The Sims Online because Linden Labs gave residents the ability to create their own content. In effect this means that the majority of the content of this world is built and owned by its users. There are some important exceptions such as the physical laws of Second Life that were created by Linden Labs, but users have control over the appearance of avatar's, the creation of objects and most elements of their physical environment.

Second Life is used for many purposes. Many educational institutions use it to simulate environments which are difficult in the actual world, for example clini-

cal settings or where students need to learn how to manage hazardous substances. Global businesses use it for meetings because of its ability to provide a simulated conference environment. However, it is also used for purposes that many would consider immoral. Prostitution occurs in Second Life, as do killings, rape, and slavery (Ludlow and Wallace 2007). Needless to say, there are morally relevant differences between the actual and virtual world instantiation of these things but their presence, even within the confines of a computer generated environment is morally debatable at least. Do virtual worlds such as Second Life cause agents to act in ways analogous to the Ring of Gyges and if so does this mean that people are only moral because of external sanctions?

Reliable general data on moral conduct in virtual worlds such as Second Life is not available. However, there is some anecdotal evidence that misbehavior in virtual worlds is more common in anonymous (non-identifiable, non-accountable) participants compared to those that are identifiable (Suler and Phillips 1998). As with anonymity, accountability for the acts of avatars within virtual worlds is variable. In Second Life there is a set of 'Community Standards' that provides guidance on what constitutes objectionable behavior of an avatar, which includes intolerance, harassment, and assault (Linden Labs 2016b). There are sanctions for violations of these standards, such as suspension of the account or expulsion from the Second Life community.

However, it is easy for another account to be created by the controller of the avatar, so the force of even the most extreme sanction may be fairly light. An example of this was described by Julian Dibbell in his article 'A Rape in CyberSpace' (Dibbell 1994). In this case, an online character ('Mr. Bungle') raped other characters in an online world, leading to calls for sanctions and the eventual elimination of Mr. Bungle from the online world (a case of virtual killing) by one of its users. It is alleged that the anonymous person controlling Mr. Bungle later returned to the community with a character named Dr. Jest.

While Dibbell's case showed that individuals at least sometimes *regard* the acts occurring in virtual worlds as morally significant, it is not clear that this view is correct, given the differences between virtual worlds and real life. We turn now to a skeptical objection that can be levelled against the view that moral conduct is possible in virtual worlds.

## CAN PEOPLE ACT IN A VIRTUAL WORLD?

While Second Life can offer anonymity, in order for it to be analogous to Glaucon's ring it must be possible for people to act in a morally relevant way. Kant shows that a moral duty necessarily implies the ability to perform the corresponding action (Kant 1998). This is partly because of the connection between obligation and moral responsibility. We can be held to account for moral obligations that we do not fulfill and this implies that we could have in fact acted on that duty. It is also important because, as Kant points out, moral responsibility implies that our will is causally efficacious: unless a moral reason can have an effect in the world it makes no sense to talk of moral action.

Avatars and the world of Second Life are virtual. The world, its objects, avatars and computer-based images are housed on the servers of Linden Labs, but viewed and controlled by actual persons sitting at keyboards in front of computer monitors. All that actual persons can do is control visual representations via a mouse and keyboard and type lines of text that other persons can read, or if they chose, speak with their actual voice. Second Life is a virtual world and not physically realized in the same way as the actual world. By contrast with physical laws that describe the real world, physical laws that govern a virtual world are commands expressed in computer code by programmers or users and resemblances to the real world can vary depending on their purposes.

Given that avatars are controlled by persons and persons can express their will via an avatar, one aspect of the Kantian precondition for moral action can be fulfilled: an avatar can express the will of a person. However, even if we accept that the will can act upon a virtual world it's not clear that its effects upon a virtual world are morally relevant, as is expressed in the following syllogism.

1.  Moral action demands that the will be causally efficacious.

2.  Even though the will can be expressed in a virtual world, it can only act upon a virtual world.

3.  In a virtual world there is no morally relevant causation.

4.  Therefore there can be no moral action in a virtual world.

Premise 3 is contentious. Its truth depends upon the kinds of actions and effects that are possible in a virtual world. In the next three sections we will discuss three actions that are usually immoral in the actual world when performed by human beings. In these cases the effects of these actions are radically different in a virtual world, in ways that affect their moral appraisal.

## VIRTUAL KILLING

Although there are disputes about the badness of death and the correct account of the morality of killing (McMahan 2002), ending a human life is ordinarily one of the most immoral things that can be done. In a situation where a person's continued existence will produce no value for them and only intolerable suffering then there are good reasons for thinking that consensual ending of that life might be permissible. Likewise in war, there are situations where there are sound moral reasons for thinking that killing might be justifiable in some circumstances.

Of course if an avatar is killed in a virtual world, ordinarily, no person actually dies. It is also pertinent that many user-created environments within Second Life and virtual worlds such as World of Warcraft are combat-based and the possibility of avatars being hurt and killed is essential to this gameplay. These facts might be considered sufficient for denying that killing in virtual worlds resembles killing in the real world in any morally relevant respect. However, the issue is more complicated because there are different kinds of virtual death. They range from a role play death where an avatar might describe their own death with words, through to the permanent deletion of the avatar from that virtual world.

A role play death has few if any future effects upon the avatar or the person controlling it. Given that role play and many combat game deaths occur within the context of a game of sorts and that the avatar is not actually killed (they merely cease to play a role in that episode of game play) it isn't appropriate to describe them as deaths at all. They are 'pretence' deaths and are similar to the pretend killings that occur in children's games. After a few seconds of lying on the floor pretending to be dead, the deceased jumps to their feet and gets on with the next game.

The most common way that an avatar dies in the sense of being deleted is via suicide, i.e. the person controlling the avatar requests that their account be deleted. Then Linden Labs will delete that avatar and they will cease to exist in Second Life. A feature of virtual suicide and virtual pretence deaths is that they are consensual. In

the case of suicide, the death of an avatar is consented to, and may be caused by, the person who has ownership of the avatar. Getting killed in a game may be a set-back to one's interests in continued play within that episode of play, and the player may strive to avoid it, but in consenting to play the game, one is consenting to the possibility of being killed and receiving this set-back. While a player in combat based game or role-play might not want their character to be killed, this is a possibility that was known before the game started. We can view such deaths as falling within what Huizinga calls 'the magic circle', meaning that because they are conduct that falls within the formal or informal rules governing that game-play, a norm has not been violated (Consalvo 2009).

Non-consensual virtual killings are possible too. We mentioned earlier the deletion of Mr. Bungle as a punitive sanction against his conduct, which included instances of virtual rape (which we discuss next). Another prominent case of this is an avatar in The Sims Online (TSO) controlled by Peter Ludlow, professor of philosophy at Toronto. His avatar edited a virtual tabloid called the Alphaville Herald in TSO and it exposed a seedy underbelly to what was supposed to be a G-rated virtual world (Ludlow and Wallace 2007). He published articles detailing how teenagers below the age of consent were providing virtual escorting services and confidence tricksters would coerce other avatars into handing over virtual property, which often had a significant actual world financial value. Ludlow attempted to log into TSO and found that the company who owned this virtual world had killed his avatar and deleted much of his property within the virtual world (*Ibid* pp 5-7).

It is, of course, absurd to think that Ludlow's avatar was in fact harmed in any morally relevant sense. Nonetheless it did harm Ludlow because he lost property, the project that he had developed, and an avenue for play and self-expression. This was a setback to his interests, and therefore a harm. The setback resembles one plausible account of the badness of death, the deprivation account. The deprivation account holds that the badness of death consists in the deprivation for a person of their future existence and the positive value this holds for them (hence consensually ending a life that holds no positive value for the person living it is not bad, and may be permissible). To the extent that Ludlow's future experiences derived from this virtual life held value, it was bad for him that he was deprived of this through the killing of his avatar. Since this act of virtual killing was non-consensual, it seems to be a prima facie case of wrongful virtual killing by the owners of TSO.

However, Peter Ludlow is still alive and has moved to Second Life where he

has created a new virtual tabloid and avatar. At least by the lights of the deprivation account of killing, the badness of the wrongful act of virtual killing depends on the degree of deprivation, or setback to interests that it causes. This means that virtual killing in the case of Ludlow's avatar was massively less of a deprivation and therefore less bad for Ludlow than killing in the real world. Despite some resemblances between the two acts, even though the killing of Ludlow's avatar is wrong it is not wrong straightforwardly for the same reasons the killing an actual person usually is. Virtual killings, even those that fall outside of 'the magic circle', do not justify the same moral obligations to not kill in a virtual world.

## VIRTUAL RAPE

Rape is unwanted, non-consensual sexual activity. While it *might* be possible to generate some cases where actual world rape is, all things considered, justified these would be rarified cases and rape is an action that is almost always wrong. As might be expected, given the findings of Malamuth and subsequent social psychologists (Malamuth 1981), virtual rape is common in Second Life. It has also occurred in other virtual worlds, such as the case of virtual rape described by Julian Dibbell (Dibbell 1994). Unlike in the actual world, rape appears to have no physical consequences in a virtual world and this might be taken to imply that it is not wrong.

1.  Virtual rape has no harmful physical consequences for the person raped (no pain or permanent bodily damage).

2.  The only morally relevant features of rape (virtual or actual) are its harmful physical consequences.

3.  Virtual rape lacks morally relevant features.

The weakness in this argument is premise two. While there is no doubt that the physical consequences of actual rape can be appalling, the psychological implications of being compelled to perform a sexual act for another person are at least as significant for its wrongness. While persons are not physically realized in a virtual world, the extent to which many identify with their avatars means that we should be more

cautious about the possible psychological effects of virtual rape, especially for those deeply attached to, or identifying with, their avatars.[1]

Moreover, we can consider whether real-world rape that caused no harm (physical or psychological) would be wrong. Such a case was imagined by John Gardner and Steven Shute:

> It is possible, although unusual, for a rapist to do no harm. A victim may be forever oblivious to the fact that she was raped, if, say, she was drugged or drunk to the point of unconsciousness when the rape was committed, and the rapist wore a condom (Gardner and Shute 2007).

Gardner and Shute defend this possibility in careful detail in their article, and argue that it is the central case of rape, separated as it is from other features that can accompany it, such as harms of a physical or psychological nature - hence they term it the case of 'pure rape'. They argue that pure rape is wrong, and cases where harms are caused along with it aggravate this central wrong. The wrong, they claim, is the sheer (i.e. non-consensual) use of a person. Using Kantian reasoning, they argue that this use is wrong not because it violates the victim's right to control bodily property, but because it denies the personhood of the victim. It does this by treating the victim as a mere source of use-value through her body.

To the extent that one identifies oneself and one's body as extending into the avatar within a virtual world, sheer use of this avatar may amount to pure rape. Moreover, it is likely that the extent to which one has this attachment will increase the likelihood of harmful effects of this virtual act on the person identifying with the avatar, aggravating the central wrong. However, this is entirely contingent upon the psychological relation between a person and their avatar, and so it is not possible to rule that all acts resembling virtual rape are instances of sheer use of a person and therefore akin to rape in the real world.

However, if all interactions in virtual worlds are consensual, this may mean that virtual rape might not be rape at all.

1.    Rape is unwanted, non-consensual sexual activity.

---

1.    Wolfendale refers to identification with one's avatar as 'avatar attachment'. For more on this, see Jessica Wolfendale (2006) "My avatar, my self: virtual harm and attachment." *Ethics and Information Technology* 9(2): 111-119.

2.  It is always possible to close the Second Life program, turn off the computer, teleport away or simply refuse to enter a virtual world such as Second Life.

3.  Actions that model rape in a virtual world must be consensual.

4.  Actions that model rape in a virtual world cannot be rape.

This argument stretches consent within virtual worlds beyond plausibility. It is hardly constitutive of life in virtual worlds that one's avatar is subject to virtual rape, so consent to join a virtual world cannot in itself constitute or imply consent to that eventuality. Even if rape is not an eventuality but a foreseeable possibility, consent given this risk is not consent to the act any more than consent to attend any situation in which sexual assault is a possibility is consent to that occurring.[2] Moreover even if 4 is correct there still could be moral reasons why actions that model rape in a virtual world (pretence rape) should not occur, but, just as was the case with virtual killings they are very different reasons from those that make actual rape wrong.

## VIRTUAL SLAVERY

Slavery is even more common in virtual worlds than pretence killing and rape. The slave-based science fiction world 'Gor', as developed in the science fiction novels of John Norman, has been realized in dozens of user-created environments within Second Life. Male and female slaves are captured, bought and sold, and used for whatever purposes their owners see fit. Still, there are some very clear differences between actual and virtual slavery that map the differences between actual and virtual rape. Virtual slavery does not harm physically and may be consensual, which calls into doubt whether virtual slavery can be considered slavery at all.

Even though virtual killing, rape and slavery may lack the effects that make their actual world counterparts so troubling this does not imply that they have no other moral significance. It might be that pretence rape increases the likelihood of an actual world rape by suggesting that because some agree to pretence rape there this is an actual world want for this too. While slave-based worlds such as Gor do have male slaves, they are very patriarchal societies where men own, rule and use slaves who are primarily female. Again, it might be that pretence slavery spills over to the actual

2.  We are grateful to an anonymous reviewer for pressing this objection.

world in some way and the attitudes and presumptions about what it is that women want influence actual world behaviour.

Virtual killing, rape and slavery are not causally connected to a kind of world where they are as morally significant as in the actual world, although it would be a mistake to dismiss their wrongfulness as insignificant. On the other hand there are other actions that avatars and the persons controlling them can perform in virtual world that are much closer to their actual world counterparts.

## VIRTUAL VERACITY

The persons controlling avatars can use them to communicate with other avatars. Typed text is often shared within the context of role play or a game of another variety. Given that these contexts involve pretence, statements that take the form of propositions shouldn't be taken as literal assertions. We shouldn't attach any great claim to truth of the child playing Monopoly's assertion that another player owes her $200,000 than we should to a roleplaying avatar's assertion that they are feeling frightened.

However, avatars type text in many other contexts. These might be statements about the actual world, such as where the person controlling the avatar lives or which time zone they are in. But, it is also possible to make many assertions that are about the virtual world, such as the price of virtual goods, the location of a virtual shop.

Whereas killing, rape and slavery are rightly thought of as 'virtual' in a virtual world when it comes to veracity there is nothing virtual or simulated about it. If we type something false with the intention that another believes it to be true it's lying. Deliberately deceiving another in a virtual world is no less real than lying in the actual world. It's a misnomer to speak of virtual veracity.

## VIRTUAL PROMISE KEEPING

The possibility of making true and false assertions implies that it is also possible to make true and false promises. These could be promises about repaying a loan of virtual money, Linden dollars, in the case of Second Life. The same reasons why it is wrong to make a lying promise in the actual world would apply to a virtual world. In both worlds the liar exploits the trust of the person deceived for financial gain. The liar disregards the moral status of the person deceived, the possibility that they might

have had plans for the money that has been loaned and uses them as a mere instrument for their purposes.

The permissibility of actions varies greatly in a virtual world with some that should almost never occur in the actual world becoming benign and others retaining the same effects. Given that the effects of actions differ so markedly we should expect the moral requirements of a virtual world to differ too. Does the difference between the actions possible in virtual and possible worlds mean that we cannot use the anonymity of virtual worlds as a test bed for the challenge to Socrates?

## VIRTUAL WORLD ACTIONS AND THE RING OF GYGES

The immorality that Glaucon predicts is similar to many of the violent virtual behaviours that occur frequently in Second Life. However because these actions can have radically different moral qualities in a virtual word we can't infer that people performing virtual forms of acts that would be wrong in the real world are acting wrongly, or at least wrongly in the same way, or that they have abandoned the internal constraints of morality.

On other hand, it is clearly possible to setback the interests of people in the real world through virtual acts (examples we have considered include virtual killing and virtual rape), and the extent to which this occurs or people make false assertions and break promises in the absence of external sanctions might provide evidence for or against Glaucon's prediction. Given the anonymity of Second Life, people can make whatever fanciful claim they wish about their abilities or status, in either actual of virtual world. As Ludlow describes there is no shortage of swindlers and con artists who will take every opportunity they get to cheat others out of virtual cash.

It might be objected that while people can hide behind an avatar, the avatar itself has an identity and a reputation. This is true to an extent but it is very easy in a virtual world to reinvent oneself, to reappear with a different identity and appearance. This is the reason we argue that virtual death is less bad than actual death. So, unlike the actual world where damage to reputation can be lifelong, a new life in a virtual world is so easy to create that the external sanctions can be trivial depending upon the investment made in that virtual life.

From even the briefest foray into virtual world it will be obvious that there are those who chose to use their anonymity in deceptive ways. It will also be obvious that there are just as many who see that veracity and promise keeping are just as important

in a virtual world. Of course it is impossible to say whether the proportion of those who act morally is any different in a virtual world, but Glaucon's claim that there would be no difference in behavior between the previously immoral and moral isn't correct.

Hume's sensible knave makes exceptions to general moral rules when it is to his advantage. Hume describes

> *... the frequent satisfaction of seeing knaves, with all their pretended cunning and abilities, betrayed by their own maxims; and while they purpose to cheat with moderation and secrecy, a tempting incident occurs, nature is frail, and they give into the snare; whence they can never extricate themselves, without a total loss of reputation, and the forfeiture of all future trust and confidence with mankind (Hume 1946, p. 155).*

While it is easy for someone to create another identity in a virtual world, if this becomes necessary because of a deception, that particular identity will forfeit trust with other persons. But, the anonymity of a virtual world will mean that this reason is not significant for any but those who have extended business networks or friendships. The second reason that Hume thinks the sensible knave misses is more relevant to why many are moral within a virtual world.

> *But were they ever so secret and successful, the honest man, if he has any tincture of philosophy, or even common observation and reflection, will discover that they themselves are, in the end, the greatest dupes, and have sacrificed the invaluable enjoyment of a character, with themselves as least, for the acquisition of worthless toys and gewgaws. How little is requisite to supply the necessities of nature? And in a view to pleasure, what comparison between the unbought satisfaction of conversation, society, study, even health and the common beauties of nature, but above all the peaceful reflection on one's own conduct: What comparison, I say, between these and the feverish, empty amusements of luxury and expence? (Hume 1946, p. 156).*

Those who cheat and swindle in virtual worlds have given up the pleasures of virtue for virtual gewgaws. Veracity and promise keeping are not virtual, even when they occur within a virtual world. Understanding the possible real-world harms, not

to mention benefits, from some virtual acts, and recognizing these as giving rise to normative reasons is morally significant. Objects in Second Life and other virtual worlds vary significantly in their value. While some of them can acquire some monetary and other value they may be viewed as 'gew gaws' in Hume's sense: feverish, empty amusements that can arouse temptation and self-interested pursuit that conflicts with virtue.

Virtual actions that are morally different from their physical world counterparts might be achieved via actions such as being deceptive and in such cases virtue has been compromised for the sake of something that is of comparatively little value. So, Hume's observation has as much or more relevance for the behavior of those who choose to do wrong in a virtual world. His answer to the challenge to Socrates is as convincing a rebuttal to those who think virtual worlds can only foster immoral behavior as it is to the general observation that we only ever appear to do the right thing because of our fear of the external sanctions of morality.

*Acknowledgements*

## REFERENCES

Brown, M. (2005). ""Setting the conditions" for Abu Ghraib: The Prison Nation Abroad." *American Quarterly 57*(3): 973-997.

Consalvo, M. (2009). "There is no magic circle." *Games and Culture 4*(4): 408-417.

Dibbell, J. (1994). "A rape in cyberspace or how an evil clown, haitian trickster spirit, two wizards and cast of dozens turned a database into a society." *Annual Survey of American Law 490*.

Gardner, J. and S. Shute (2007). The wrongness of rape. Oxford, Oxford University Press.

Gillett, G. and C. Pigden (1996). "Milgram, method and morality." *Journal of Applied Philosophy 13*(3): 233-250.

Glover, J. (2000). Humanity: a moral history of the twentieth century. New Haven, Yale University Press.

Hume, D. (1946). An enquiry concerning human understanding and selections from a treatise of human nature. Chicago, Open Court Publishers.

Kant, I. (1998). Groundwork of the metaphysics of morals. Cambridge, Cambridge University.

Linden Labs, (2016a). "Build worlds with us." from http://www.lindenlab.com/releases/linden-lab-invites-first-virtual-experience-creators-to-project-sansar-testing [Accessed November 2017].

———(2016b). "Community standards." from https://secondlife.com/corporate/cs.php [Accessed Novemer 2017].

Levy, K. (2014). Second Life has devolved into a post-apocalyptic virtual world, and the weirdest thing is how many people still use it. *Business Insider*.

Ludlow, P. and M. Wallace (2007). The Second Life Herald: the virtual tabloid that witnessed the dawn of the metaverse. Cambridge, Mass., MIT Press.

Malamuth, N. (1981). "Rape proclivity among males." *Journal of Social Issues* 37(4): 138-157.

McMahan, J. (2002). The ethics of killing: problems at the margins of life. New York, Oxford University Press.

Milgram, S. (1963). "Behavioural study of obedience." *The Journal of Abnormal and Social Psychology* 67(4): 371-378.

Nogami, T. and J. Takai (2008). "Effects of anonymity on antisocial behaviour committee by individuals." *Psychological Reports* 102(1): 119-130.

Plato (1993). The Republic. London, Pimlico.

Suler, J. and W. Phillips (1998). "The bad boys of cyberspace: deviant behaviour in a multimedia chat community." *Cyberpsychology and Behaviour* 1(3): 275-294.

Wolfendale, J. (2006). "My avatar, my self: virtual harm and attachment." *Ethics and Information Technology* 9(2): 111-119.

# Population and Having Children Now

JAN NARVESON

*University of Waterloo*

## ABSTRACT

This paper aims to state the obvious - the commonsense, rational approach to child-producing. We have no general obligation to promote either the "general happiness" or the equalization of this and that. We have children if we want them, if their life prospects are decent - and if we can afford them, which is a considerable part of their life prospects being OK - and provided that in doing so we do not inflict injury on others. It's extremely difficult to do this latter, but affording them, in rich countries, is another matter. With that qualification, by and large people should just go ahead and have (or not have) children - as many as they think they want and can handle - as it suits them.

Over many years I wrote several times about population. In my first paper on the subject "Utilitarianism and New Generations" (1967a) I was interested in exploring utilitarianism, or so I thought. My main point was that utilitarianism did not commit one to making happy people, but only to making people happy. Something like that continued to be the theme in two subsequent papers (Narveson, 1973; 1978). Meanwhile, Derek Parfit came along, with his interesting questions about what happens in cases of nonidentity—where the children are not the same people that would have existed if some other decision had been made or policy imposed (Parfit, 1983). So I feel motivated to rejoin the issue, in the wake of having, several decades ago now, abandoned my original proclivity toward utilitarianism. All these matters need restating and reinvestigating.

So let's start where I left off. Should we say that in having children, you *benefit them* (if they are happy), and possibly you *damage* or *wrong* them (if they are miserable)? My original "insight" (if that's what it was) remains unchanged: No. You don't

*confer* life on someone when you bring them into existence, because there isn't any "one" upon which to confer "life." People's lives aren't something else, superadded on top of "them." By the time they are born, there's no "conferring" left to do, and prior to then, there is no one on whom to "confer" anything.

A different issue would be that of, say, abortion, where there *is* a "something"— viz., an embryo or fetus—and abortion would, of course, destroy the life of that thing, the only question then being whether what you thereby take the life of is a morally countable person—a morally human person and thus a possessor of rights including the right not to be killed. We aren't discussing that topic in this inquiry, however: we are addressing, simply, procreation. Whether you count as "procreating" when you produce an embryo that is not carried to term is separate from the present issue, which is, merely, why and whether you ought to procreate at all, whatever we decide procreation precisely consists of.

In procreating, people aim—if they aim—to bring new people into existence. To be sure, they may not "aim"—they may well conceive unintentionally, or contrary to their intentions. But our question is philosophical: what are good reasons to have children, and what are good reasons not to, whatever may have actuated some particular couple at a given time? In particular, we want to know whether, if children we would have would be happy (or whatever our favored terms are: would thrive, flourish, do well) *and* we can know this, then *that* is, as it stands, a good reason to have them.

Regarding that question: Contrary, perhaps, to what I have been (perhaps mis) understood to say in those earlier papers, the correct answer is: of course! The existence of people leading good lives is (normally) a good thing. So far, so good. But now: is it a good thing because of what it does *to* those people—as would be the case with a benefit conferred on some already-existing person, such as a victim of disaster in some faraway place? No, and for the same root reason: there is no *person* upon whom life is conferred when we have children. If anything can be said to be such that creating a life means "conferring" something on *it*, it would have to be some such thing as a bit of germ plasm. And surely no one thinks we have moral duties toward those things, as such. Our duties, always and by definition, are to *people*; any sort of obligation to any sort of things other than people must be indirect, a consequence of some commitment to some person(s) or other. (Moreover, we will take it here that all duties are to *other* people, apart from duties "to" oneself that derive from commitments made or duties held towards others.) Having good people is, simply, a good: it

is also, usually, a "public" good—we'll say more about this below. And by definition, the goodness of the life thus produced is a good *to its possessor*. It may be a good for others too, of course—or perhaps the reverse. But our question here is whether the fact that people are brought into existence and lead happy lives is, in and of itself, to be regarded as good, and thus as a reason, so far as it goes, for so bringing them. And that's the question I'm answering in the affirmative.

Of course, the birth of a child with good prospects is a direct benefit to its parents, assuming they have normal feelings for their family, and similarly to any other well-wishers relevant to the case. Moreover, it is also, with any luck, a benefit to humanity at large. One hardly needs to enumerate. It isn't just that the child might be exceptionally talented and make audiences, or colleagues, happy to see its good performances or to read its important results etc. But perfectly ordinary people of no special talents are also a good to many people around them—again, with any luck. (I specifically include people born in what are by world standards very (economically) poor circumstances. Only those who, for instance, become diseased early on, or who otherwise do exceptionally badly, should be regarded as tragedies. And even they have often contributed to the pleasure and general benefit of many people meanwhile. Whatever the grown child does economically is also a benefit, to all those who use his or her products.

In saying this, I do *not* mean to re-embrace utilitarianism. Our question is: why should the goodness of good lives matter to *us*? That is an important question, not to be sloughed off. There are just two sorts of answers. (1) If we are, say, its parents or simply benevolent persons, they "just do" matter to us—we, those of us falling under such descriptions, *like it* that there are happy people out there (and of course, this will, alas, not always be so; there are those who want to abandon their newborns, and many who wish they had not become pregnant in the first place); and (2) much more importantly so far as social philosophy is concerned, more good people are a benefit to the other people in the social environment, local or global, into which they are born, whatever anyone's personal attitudes may be. If this latter were not so, then there would be room for concern, and possibly for imposing constraints. But that it generally is so, at very least since a couple of centuries ago, is, I believe, clear—even

though many philosophers and others seem to have persuaded themselves, despite all evidence, that it is not so.[1] Again, we will address this further below.

There has been a radical challenge to the claim that producing people with good lives is of value, sufficient to justify bringing them into existence. David Benatar (2008) argues that it is always better that a person not exist. The basis of this counter-intuitive view is that harms can be avoided by nonexistence—which is true enough, of course. Benatar insists, "Those who never exist cannot be deprived. However, by coming into existence one does suffer quite serious harms that could not have be-fallen one had one not come into existence." Also true. But so what? If the good out-weighs the bad in a life, why isn't that good enough? It is true that we ought not to harm others. But bringing people into existence who will, no doubt, be sometimes harmed, is not *harming them* in the sense in which that is a morally wrong thing to do. It is creating something that becomes a person, who can and very likely on occasion will be harmed. But you (the parents) *didn't do it*. (Not normally, that is. When fore-seeably a baby would be born with some horrible disease which will also kill it soon, the case is different.) Moreover, in normal interpersonal cases, it is quite possible for harms to be outweighed by benefits, as when the dentist's infliction of pain leads to well-functioning teeth, which in turn affords nutritional and culinary benefits that, in the view of the patron of the dentist, well justifies that modest infliction. Late in life, too, we often experience great pain and enormous inconvenience, but still prefer life to death. Why wouldn't Benatar think it worthwhile to give people that choice? I don't see a good answer to that.

To return to our main issue: the question now is whether the fact that one could relieve the suffering of others by not having children of our own, as perhaps some-times is the case, implies that we ought not to have those children. Does it follow, especially, from the fact (when it is a fact) that we could devote resources to the im-provement of conditions for presently-suffering people, and specifically the resourc-es that we will otherwise devote to our hypothetical happy child once born, that we therefore *ought not* to have our own children instead? Might this be on argued the ground, say, that alleviating misery is more important than increasing the goodness of good lives?

To this last that I also want to say No. Specifically, I want to advance the view,

---

1.    An instructive example: Paul and Anne Ehrlich, The Population Bomb (1968)—a book whose refutation at the hands of reality was well under way at the very time it was written, and overwhelm-ingly reiterated since.

which I take to be common sense, that people are entitled to have children so long as the effect of having those children is not to *impose* significant costs, harms, disbenefits, on other people[2]. It can pass that test, however, without thereby improving the lot of suffering people, wherever they may be, whom we might perhaps have otherwise been able to benefit if we so chose. Producing our own children does not (normally) harm others, though it conceivably might leave them worse off than if one instead had devoted resources to benefiting them. But that is not the same as harming them. This, then, is an outright rejection of utilitarianism, and certainly of "prioritarianism," a view that has had considerable prominence of recent times[3] and to which I was myself partial in my earlier years (see Narveson, 1967b).

Is this changed when we take into account the interests of the person produced? If, as I have asserted, we cannot be said to harm someone by bringing him or her "into existence," then have we no responsibility to produce people whose lives will be better rather than worse? Once born, of course, we have massive obligations to the child, who is extremely vulnerable in early years, and whose later character is so substantially influenced by what is done to that child in earlier years. As Weinberg has it, having children is "hazardous": the potential for producing severely damaged, deprived, or otherwise miserable children is real (Weinberg, 2016). But prior to their birth, can anticipation of bad results yield a block against child production?

Here again, I come down on the No side, and for the same reason. When you produce a child, there is creation *ex nihilo*! That is: a person somehow emerges from the protoplasm resulting from the combination of sperm and egg (or, from the very late-term foetus—again, to distance ourselves from the dispute about the proper stage of organic development to identify with the onset of genuine personhood.) Prior to the act by which those got together, there was no such person. But the notions of harm and benefit only apply to persons. A is better or worse off at t2 than A at t1. But if there is *no* person at t1, we have lost our comparator. We—parents, onlookers, philosophers, and so on—can make judgments as to the level of well-being enjoyed by this person, or suffered by it. But so can the new person himself. We can try also to compare how he is with how others are, or with the different individual we might

2.   "Other people" being actual, identifiable persons or groups of such—not "possible people," simply as such. Possibly some of those persons are in the future, to be sure—questions about that are discussed in the body of this paper.

3.   E.g., the popular philosopher Peter Singer, whose many articles to this effect simply fail to square with the beliefs, especially as implied by their actions, of ordinary people.

have conceived instead, if the luck of the draw, or planning, or impinging circumstances, had been better.

Now: *some* individuals so produced will prove very costly to maintain in life, or in any condition that its parents, or others, would like to see it in. And about this we can certainly moralize. The view here is that, fundamentally, these costs are to be borne *by the parent(s)*—that's the measure of the "hazards" Weinberg intends. If there is no way to internalize those costs, then those parents have acted wrongly or at least irresponsibly. The rest of us had no obligation to bear those costs and if we do anyway, that's nice of us—but it is *not* a duty (normally) so to bear them.

Your experience of parenthood may well be better had you produced one of the other possible persons than the one you actually did. If so, that's a reason to think ahead, insofar as possible, and to have children you know you can afford and that you are sure you can love and rear in joy. (And the parental love you bestow on that child will pay dividends for her or him in later life, too.)

I previously said that my position is an outright rejection of utilitarianism, and also of "prioritarianism." But actually, it may not entail such a rejection, unless those views are interpreted in a certain way. To see this, we now need to make a distinction.

## TWO DEPARTMENTS OF MORALITY

Morals generally consists of a set of directives: rules, apothegms, principles of action (and attributions of "virtue" or "vice," say.) They are not the directives of anyone in particular, and certainly not someone "in authority"—there are no moral "authorities." Nor are they the deliverances of some privileged group. Notionally, they are the directives issued by just anybody and everybody—hopefully, the latter, and unanimously. Lack of unanimity presents a problem, perhaps soluble. But we'll go no further at the moment (we'll be back to it.[4]

Now, some of these directives are of the general drift, "it would be a good thing if you were to do X" or "it would be virtuous to do X", or perhaps "you get points for doing X!"; while others are of the drift, "if you don't do x, we ["society," possibly via agents acting supposedly on its behalf] will inflict certain losses or injuries on you." That is to say, some—the latter sort of—moral directives are meant to be *enforceable*

---

4.    My general views on the foundations and nature of morals are expanded in This is Ethical Theory (2009)

by punishments, while others—the first sort—are meant not to be such, but rather, *reinforceable*, as we should more mildly put it, by, roughly speaking, praise or blame and their ilk. There's a difference of tone, true—but also a difference of real consequence to the actors in question. They had, we might put it, better sit up and listen if the punishable sort are in question; the first sort, on the other hand, they could perhaps ignore, though they *shouldn't*, of course. Broadly, it's a difference of stick and carrot, both wielded, notionally, on behalf of some society in particular, or (better) of mankind at large.

So the question about population is this: When do we get to intervene to prevent people *from* having children, or alternatively, perhaps to compel them *to* have them? On what ground, if any, do we get to threaten ill consequences to those who prefer some other option than what is being proposed by the theorist or the moralist in question? To this, I think, there is a good answer, and it is the same as when any other sort of illfare-imposing actions are in question: We would get to do so if your having, or not having, the children in question would *work ill for some persons.* Not, however, for some fancied future possible persons, but for actual persons, alive, breathing, and acting *now* (or at any rate in their lifetimes.)

On the other hand, we do *not* get to do so just because the future progeny whose production or nonproduction is being contemplated by the agents being addressed, would be happy or not. And that isn't just because somehow this pattern of reproductive activity does not accord with some calculated best state of mankind in future. If, of course, the person in question likes that calculated idea and wants to conform his behavior to it, no problem. But that's trivial. What isn't trivial is when someone does *not* like the idea and wants to have her or his own children, or not, for her or his own reasons. As, I would think, somewhere around 99% of mankind do.

Why do they (we) have children? Because our (would-be grand-) parents *really* want us to have them; because we just love the idea of having children around, or love children themselves; because we want to continue our particular DNA type; because we hope they will help support us in old age; because we didn't know of contraception technology, or the one we usually use failed on this occasion; because we think the community needs more people of the type we expect our children would be; or..... and so on. Now, the question posed by philosophers of the at-least-partially utilitarian stripe is: shouldn't we be using our resources to help out the suffering, or the less-fortunate, or the oppressed, etc., among us first? There are two possible lines of response to this.

One is: No. It's tough that they are in such bad shape, but it's not my fault that they are, and my interest in having my *own* children outweighs any interest I have in helping out those people. (As it might not. I might be highly sympathetic to those people. But the point is, I might not be—and I daresay most people most of the time are not, or at least, certainly not enough to defer or renounce the having of their own children instead. Or, as may be, the building of that summer cottage they always wanted.)

Another is: well, maybe I "should"; but, frankly, to us (my partner and I, say) having our own children is *more important*. More important to us, that is. But then, *we are* "us" in this case, and because we are, our values are determinative of the matter, so long as we do not actively clash with actual others thereby. Our values are what move us to do what we do—*not* the values of some or a lot of actual or possible or other persons, future or present.

Let us now imagine that the persons contemplating reproduction (which, very likely, they don't—but, as mentioned above, we are asking what there is to say for given behavior whether or not the agent of that behavior actually says it or not) are members of a marginal group. They live perhaps by hunting and gathering, or marginal farming, and they are more than ordinarily subject to debilitating diseases, and so on. That is to say, let's consider these questions as posed by the very persons figuring in Parfit's "repugnant conclusion"—people living the kind of lives we (we philosophers, perhaps, or we ordinary people) would hope and prefer that nobody would have to live. We may conjecture that those people aren't about to stop having children just because they'll be poor. Why, then, should other people who are better off stop having them in order to perhaps enable some of those poorer people to be less poor?

## ALLEGED "RESOURCE SCARCITIES"

At this juncture, to be sure, we will bump up against familiar objections concerning resources. What, it will be asked, if there just isn't enough to go around? One cause of war, no doubt, is competition over resources—real or, by far mostly, imagined. The Nazis claimed that they needed "lebensraum" and would have to get it from the Slavic peoples to their east. Whether any such claims were ever true in history is probably an interesting question, but it is certain that they are not true anymore and were not true in Hitler's time. The many billions of us on earth now are all able

to live far better lives than our remote ancestors because of the developed productive powers of humankind. Scarcity of the general kind posited by Malthus simply isn't on any more, if it ever was.[5] And, further, production is stimulated primarily by the prospect of improved income from exchange by the producers. That certainly includes exchange with the erstwhile indigent, such as the Chinese, many hundreds of millions of whom have advanced from severe poverty to relative affluence in the recent past because of their government's change in the direction of relatively liberal economic policies, including extensive trade with much of the rest of the world.

So the premises required for positing a necessary trade-off between better-off and worse-off are, at least in the current and foreseeable era, false. It is still true that in individual cases, people *can* choose between devoting their resources to their own children, or to relieving poverty—just as they can choose between that and supporting the opera. And as a matter of fact, their choices, all over the world, have overwhelmingly supported smaller families, though not for the philosophical reasons pressed by so many, but simply in contemplation of their own economic circumstances. But do the rest of us have any right to insist that they do the one rather than the other? No. The lives of the meanest among us are still *their* lives, and *not* ours, or some institution's. When it comes to insisting, we may indeed insist on decent behavior—normally a readily meetable standard. But beyond that, we must deal with them as with anyone else: by negotiation, with a view to advancing mutual interests. And it may not be in the *mutual* interests of some wealthy person here and those of some impoverished person over there that the former should help out the latter.

There has, especially of recent times, arisen a view that it is actually quite easy for us to impose "harms" on future generations, and one version of that sort of view even claims that we can do so simply by having children, either just at all, or at any rate with exceptions that have to meet quite stringent conditions[6]. This general view strikes me, and most people, I would think, as more than slightly incredible on the face of it—and upon deeper inspection, too. Still, it or something like it is popular, and deserves some attention.

Of course, the main driver of this idea has always been that the more there are of us, the now-people, then the less resources there will be for *them*, the yet-to-exist people. It is perhaps more difficult to get this thoroughly exploded and incoherent idea out of people's heads than any other misunderstanding about the world around

5.  Among many books on these matters, see for instance Ridley (2010) or Bailey (2015)
6.   See, for instance, Daniel Benatar (2008)

us. The true situation is just the opposite. Resources are not fixed: they are created, by us, and the more of us there are working away at it the more there will be—not the less. If we suddenly reverted to the world as it was, say, 40,000 years ago, the gross mass of the world around us would be the same, or maybe a bit greater. There'd be a lot more oil down there somewhere—oil that nobody had any idea how to use; and a bit more iron; etc. Now: would you think yourself much richer if you were one of the few projected into that world? You'd be deprived of ready energy, cars, nice houses, and on and on and on, and life would be pretty much "nasty, brutish, and short."

Those who say this seem also to be very cavalier about technology. They look to what they think to be the "perils" of global warming, and suppose that less people would mean less $CO_2$ in the environment, whereas more people means more. But apart from the specifics of all this—the still-debated reasoning behind predictions of very substantial climate change[7], for example—two things are extremely clear. First, that the methods by which we are attempting to reduce the incidence of greenhouse gases are hugely, indeed fabulously inefficient, and consequently incredibly expensive—with the expense falling squarely on the backs and pockets of the poor. No rational person would propose to "save" him or herself from a supposed danger by methods such as that. We would not, individually, pay a hundred thousand dollars a year to get our garbage picked up with 1% cleaner methods. But politicians are wonderful at talking us into paying that and more for "environmental" controls that do far less and cost far more.

Any major perils predicted—seriously higher sea levels, say, such as to endanger large cities—are going to come about, if at all, quite a long way down the road, so far as I can see. Even on this there is disagreement, to be sure. The minimum "road" seems to be several decades, though (see Nuccitelli, 2015, pp155-6) for perhaps the most extreme current view). In the meantime, very smart people are working away at ideas to deal with this, if indeed it threatens. Some of those, as time goes by, will be far, far better than anything we are doing now. What's our hurry?

As a salient example, many countries and provinces (such as my own, Ontario) are investing heavily in "clean" electric production via wind turbines and solar arrays. But both of those have a perfectly obvious, and for the near past and predictable future, overwhelming problem: they only work when the sun shines (at most half the

---

7.    The "climate change" literature is now immense, of course. For critics of the "received" view, see for example Alan Moran (2015). Defending the "received" view, see for example the characteristically titled Climatology versus Pseudoscience by Dana Nuccitelli (2015) Reading the two in tandem is strongly recommended.

time) or the wind blows (more like a quarter). So what to do during those inevitable very frequent lulls? The only feasible answer at present is to revert to the very kind of production that "clean" was supposed to avoid, of course. And so we pay for massive overcapitalization in this industry, with electricity at twice the price it could be[8]. And who pays that? We, the ordinary people, do. The poor, of course, are the chief sufferers. The main beneficiaries are these who supply the equipment and the technical support.

The general point is that insofar as arguments for population restriction appeal to a supposed conflict of interest between current people and future people, those arguments are empirically questionable, at the least. But we can go farther. For even if one *knew* that one's having a child might disenable some identifiable future person from having one, or as happy a one, it would *still* be this actual present person's right to have her own child instead. And this time, the reason goes back to the original "insight": that possible future other person doesn't exist, and so we can't be said to owe *it* anything at all. When the time comes, after all, the would-be parents of that possible child have, as we always have had, the option not to have it instead, in view of current circumstances.

Compare this with the case where your child and that of Ms. Smith are to emerge simultaneously from those respective wombs, and that there's a medicine needed to enable successful birth—but, unfortunately, not enough for both of you. What then? Various resolutions might be possible: Ms. Smith *ordered* the medicine and you did not; you offered her $100,000 if you could have it and she not, and she accepted, etc. Instead, though, let's suppose that there is indeed a straight conflict. In that case, certainly, a throw of the dice or the equivalent is called for, and neither has prior claim. Whereas if our child would be born on Wednesday and hers on Saturday following, then we *got there first*. Or, one of us had insurance giving that person first claim on the needed medicine. But the very rarity of such cases is what points up the general

---

8.     We can't get too involved in technicalia here, but readers should be aware that the "costless" "green" electrical generation technologies have the intrinsic disadvantage that the sun shines at most half the time, and the wind blows with sufficient strength (but not too much) about 26% of the time. So where do we get the needed electricity for serious grids feeding busy cities and the like, during that very large down-time? The only current really feasible answer is: from non-green sources such as gas or coal. This means that heavy utilization of the "green" technologies entails duplicate capitalization of everything and enormous supply problems. In my home Province of Ontario, which has gone in heavily for these new sources of electricity, the result has been a more than doubling of electricity prices for consumers over the past several years, and heavy deficits for the government. One result is that there is either no gain at all, or virtually no gain, in the very purpose for which these are built, namely to reduce net $CO_2$ admissions. Enthusiasts for such technologies, many of whom have profited enormously from the heavy subsidization of these industries by government, certainly ought to have thought of these problems before. For a thorough technical analysis, see Etherington (2006).

moral, which is that each person does his or her best to promote her interests, letting the chips fall where they may—and *rightly so*. A supposedly philosophical interest in the welfare of future generations evaporates in the face of reality, which is *this* person having *her* child, which she very much wants.

Perhaps it will be envisaged that there might be a sort of bargain with the current parents-to-be, or grandparents-to-be, or nth generation to-be, of the "future generations" in question, which obligates us to have less or no children, etc. And of course, there might. But again, there might not. Why would present persons really desirous of having families engage in any such bargain?

Let's remember that a majority of present potential parents have *elected* to have few or no children, as is their right. But our question is, what about the others who would like to have larger families, and who can afford it? There is, I suggest, no excuse for thinking that these people may be *required* to curtail their desires in that regard. *Only* when a real and concrete hurt can be expected to be imposed on currently known persons would there be any case for that. And that is possible—but there is nothing inevitable about it, and indeed, it is very typically not so.

The general conclusion, then, is that philosophical attempts to justify impositions on aspiring parents are seriously flawed. Common sense recommends that we pay little attention to the scenarios they want to invoke to that end. Human life can and should continue, where desired, so long as it doesn't outrightly conflict with the lives of others.

## REFERENCES

Bailey, R. (2015) *The End of Doom: Environmental Renewal in the Twenty-First Century* (St, Martin's Press, 2015)

Benatar, D. (2008) *Better Never to have Been* (Oxford University Press, 2008)

Erlich, P. and Erlich, A. (1968) *The Population Bomb* (Ballantine Books 1968)

Etherington, J. (2006) *The Wind Farm Scam* (Stacey International, 2006)

Moran, A. (ed), (2015) *Climate Change: The Facts* (Melbourne: Institute of Public Affairs: Stockade Books, 2015.)

Narveson, J. (1967a) "Utilitarianism and New Generations" *Mind*, vol. 76, No. 301, January 1967, pp. 62-72 (reprinted in: S. Gorovitz, Mill's Utilitarianism, Text & Commentary (Bobbs-Merrill, 1971)

———(1967b), Morality and Utility (Johns Hopkins Press, 1967)

———— (1973) "Moral Problems of Population", *The Monist*, vol. 57 No. 1—January 1973, pp. 62-86. Reprinted in M. Bayles, Ethics and Population (Cambridge: Schenkman, 1976)

———— (1978) "Future People and Us", R. I. Sikora and Brian Barry, eds., *Obligations to Future Generations,* Philadelphia: Temple University Press, 1978 (Reprinted in A.R. Gini, David Ozar, and Patricia Werhane, Philosophical issues in Human Rights (Random House, 1985)

———— (2009) *This is Ethical Theory* (Peru, Illinois: Open Court, 2009)

Nuccitelli, D. (2015) *Climatology versus Pseudoscience* (Santa Monica, CA: Praeger 2015)

Parfit, D. (1983) "Future Generations," in Narveson, ed., *Moral Issues* (Oxford 1983, 414-444)

Ridley, M. (2010) *The Rational Optimist: How Prosperity Evolve*s (NY: Harper Collins, 2010)

Singer, P. (1972) "Famine, Affluence, and Morality," *Philosophy and Public Affairs* 1972. (3): 229-243.

Weinberg, R. (2016) *The Risk of a Lifetime: How, When, and Why Procreation May Be Permissible,* Oxford University Press, 2016. Review of this book by David A. Jensen, Brigham Young University in Notre Dame Philosophical Reviews, May 19, 2016.

# Is Sex With Robots Rape?

ROMY ESKENS

*University of Oxford*

2017 OXFORD UEHIRO PRIZE IN PRACTICAL ETHICS

WINNER: GRADUATE CATEGORY

## ABSTRACT

It is widely accepted that valid consent is a necessary condition for permissible sexual activity. Since non-human animals, children, and individuals who are severely cognitively disabled, heavily intoxicated or unconscious, lack the cognitive capacity to give valid consent, this condition explains why it is impermissible to have sex with them. However, contrary to common intuitions, the same condition seems to render it impermissible to have sex with robots, for they too are incapable of consenting to sex due to insufficient cognitive capacitation. This paper explores whether the intuition that non-consensual sex with robots is permissible can be vindicated, whilst preserving valid consent as a general requirement for permissible sexual activity. I develop and evaluate four possible ways to argue that there is a morally significant difference between robots on the one hand, and insufficiently cognitively capacitated humans and non-human animals on the other hand, to substantiate and justify the intuition that it is permissible to have non-consensual sex with the former but not with the latter.

## 0. INTRODUCTION

In ethical discourse, valid consent is usually taken to be a necessary condition for permissible sexual activity (Archard 2007; Cahill 2001; Cowling et al. 2017; Soble et al. 2002). Since non-human animals, children, and individuals who are severely cognitively disabled, heavily intoxicated or unconscious, lack the cognitive capacity to give valid consent, this condition renders it impermissible to have sex with them. However, the claim that valid consent is necessary for permissible sexual activity seems to imply that it is also impermissible to have sex with robots, for they too are incapable of consenting because they lack the requisite cognitive capacities.

I suspect that many people find this apparent implication counterintuitive, even though they would endorse the view that valid consent is a necessary condition for permissible sex. Hence, many people seem to hold conflicting intuitions about the permissibility of non-consensual sex between cases in which the non-consenting subject is a robot and cases in which they are a human or non-human animal. To show that these intuitions are in fact consistent, so that neither has to be abandoned, it must be demonstrated that there is a morally significant difference between the two types of cases, which explains why valid consent is a necessary condition for permissible sexual activity if the non-consenting subject is a human or non-human animal but not when they are a robot. Solving this puzzle challenges us to refine our conception of the relation between consent and the permissibility of sexual activity, and to reflect on the ethics of robot-human interaction.

This essay proceeds in the following way. In Section 1, I elaborate on the current and anticipated design of sex robots, as well as the sex robot industry, and provide a brief overview of the ethical debate on sex robots. In section 2, I clarify the notion of valid consent, and expand on its application in sexual ethics. In section 3, I explore and reject three possible ways to argue that there is a moral difference between robots on the one hand, and humans and non-human animals on the other hand, which renders it impermissible to have non-consensual sex with the latter but not with the former. In section 4, I argue that this moral difference is grounded in a discrepancy in moral status. Section 5 concludes.

## 1. SEX ROBOTS

*1.1 Design, Demand, and Development*

Sex robots ('sexbots') currently make up the only category of robots capable of having sex. Sexbots are robots designed for the purpose of providing sex for humans. They are programmed to engage in sexual activity, as their governing software necessitates them to perform acts aimed at the sexual gratification of their users (Levy 2009; Sparrow 2017). The sexbots currently available for sale look and behave moderately like humans, insofar as they have silicone skin and are realistically shaped, and are equipped with a rudimentary AI system that enables them to interact with their users through speech and affective communication (Levy 2009; Sharkey et al. 2017; Sparrow 2017). The sexbot companies that are currently best known are TrueCompanion, which has launched multiple editions of the female sexbot Roxxxy, and RealDoll, which sells customizable models of both male and female sexbots.[1]

A number of surveys indicates that the market for sexbots has potential for significant expansion in the future. Szczuka and Krämer, for example, found that 40,3 percent of the 229 heterosexual males who participated in their study could imagine buying a sexbot within the next five years (2017). Likewise, Scheutz and Arnold conducted a small-scale survey, indicating that two thirds of the male participants were open to the idea of using sexbots in the near future (2016).[2] Furthermore, Li, Ju and Reeves found that participants in their study were subject to increased physical arousal when they touched the 'erotic zones' of a sexbot, which suggests that there generally is no physiological barrier to sexual relationships between humans and robots (2016).

Moreover, experts believe that it lies within the scope of near future technological possibilities to create sexbots that are far more sophisticated than those currently available. These sexbots will be autonomous and interactive, with flesh-like skin, affective computing, highly developed sensory perception, refined language skills, the capacity to learn, and multiple preprogrammed personalities. They will adapt to the

---

1.    See www.truecompanion.com and www.realdoll.com (both accessed 28 June 2017). Other sexbot companies that are currently competing on the market are Android Love Dolls and Sex Bot Company.

2.    In stark contrast to the male participants, Scheutz and Arnold found that two thirds of the female participants declared themselves to be *opposed* to the idea of having sex with a sexbot. I can not elaborate on this gender difference here. However, see Scheutz and Arnold (2016) for possible explanations of this result.

sexual preferences of their user and base their sexual performance on an extensive amount of data, which will enable them to provide sexual gratification for their users (Levy 2009; Richardson 2015; 2016; Sharkey et al. 2017; Sparrow 2017).

These sexbots of the near future will also display sentient behaviour, such as the experience of sexual pleasure when internal sensors are triggered, without actually having qualitative experiences. Moreover, their programming will cause them to demonstrate complex emotional behaviour, which will allow for the formation of intimate sexual and emotional relationships with their users. It is this characteristic that differentiates sexbots from sextoys. The creation of such sophisticated humanoid sexbots can be expected to increase the number of individuals that is open to the idea of having sex with a robot (Levy 2009; Richardson 2015; 2016; Sharkey et al. 2017; Sparrow 2017).[3]

In this essay, I will be concerned solely with near future sexbots. Hence, my conclusions might no longer hold if sexbots become increasingly advanced in the more distant future.

*1.2 Ethical Considerations*

David Levy has argued that there are strong reasons in favour of creating sophisticated sexbots (2009). To begin with, he suggests that sexbots could replace human sex workers, which has the potential to curtail persisting harmful practices in the sex industry, such as sex slavery and sexual abuse. Furthermore, he maintains that the availability of sexbots could offer a solution for individuals who experience difficulty finding sexual partners, and provide intimate companionship for those who feel lonely or isolated.[4] Moreover, it might be possible for sexbots to function as satisfactory alternatives for individuals with sexual desires that are likely to cause harm

---

3.    This anticipation is also supported by the results from a 2016 survey conducted by Nesta. See Nesta FutureFest Survey, April 27, 2016, http://www.comresglobal.com/polls/nesta-futurefest-survey-2/ (accessed 23-07-2017).

4.    For objections to this claim, see Giutu (2012) and Sullins (2012).

to humans if brought into practice, such as the desire to have sex with children or to engage in violent or degrading sex without another agent's consent.[5]

There are also legitimate worries about possible negative effects of the development of the sexbot industry. Kathleen Richardson, who launched a Campaign Against Sex Robots in 2015, has stressed that the sexbot industry is mostly driven by heterosexual men who desire to buy robots that look realistically like women, but appear and behave in ways that exhibit harmful pornographic stereotypes of female attractiveness, personality and sexuality (2015; 2016). According to Richardson, this fact causes the design, sale and use of sexbots to express and reinforce persistent sexist and misogynist stereotypes about females, which she expects to have harmful consequences for women (2015; 2016; see also Scheutz and Arnold 2016; Szczuka and Krämer 2017).[6]

Even if the design, sale and use of sexbots will have no such effect on women, it might be wrong for individuals to interact with sexbots in ways that would be harmful if the recipients had been actual women. Robert Sparrow has argued that behaviour that has no extrinsic harmful effects can be morally objectionable in virtue of *representing* harmful behaviour (2017). Indeed, he maintains that because sexbots are intended to represent women, sexist and misogynistic behaviour towards sexbots is morally objectionable in virtue of representing such behaviour towards women (Sparrow 2017). Another way in which sexist or misogynist behaviour towards sexbots can be considered morally objectionable without having harmful effects on others concerns moral character. According to John Danaher, for instance, such behaviour is expressive of the defective moral character of the sexbot user, and is to be condemned on that basis (2017; see also Sparrow 2017).

There are thus good reasons to be cautious of the design, sale and use of sexbots, although these reasons are far from definite. However, since the subject of this essay is the permissibility of non-consensual sexual relationships with robots, I will set these issues aside for the remainder of this essay. Nonetheless, it should be kept in

---

5.    This claim is highly disputed both within and outside of the academic community. This is in part because there is insufficient empirical evidence to support or oppose the claim that the desire to have potentially harmful sex with actual people diminishes when substitutes, such as sexbots, sex-dolls and virtual pornography, are used. It is unlikely that such evidence will become available soon, since research on the topic is suffering from a lack of funding and a scarcity of willing participants, largely due to the ethical complexities involved in such research, and the moral contentiousness of the research subject. Nonetheless, for an overview of some of the arguments for and against the use of sexbots as substitute subjects for potentially harmful sexual practices, see Rutkin (2016).

6.    See www.campaignagainstsexrobots.org/ (accessed 28 June 2017). For objections to Richardson's campaign against sexbots, see Devlin (2015).

mind that although it might not be impermissible to have sex with robots for the reason that it is non-consensual, it might be impermissible for other reasons.

## 2. VALID SEXUAL CONSENT

*2.1 When Is Consent Valid?*

There is a general consensus in ethical discourse that valid consent is a necessary condition for permissible sexual activity (see for instance Archard 2007; Cahill 2001; Cowling et al. 2017; Soble et al. 2002). There are many ongoing debates that concern the concept of valid consent or its applications to moral and legal disputes. Since this essay concentrates on clear cases of sexual non-consent, I will not engage extensively with these issues here, and this section will involve only a brief discussion of the notion of valid consent and its relevance to sexual ethics.

Most philosophers engaged in the debate on valid consent identify three independent necessary conditions that are together sufficient for consent to be valid (Kleinig 2010; Mappes 2002; Miller and Wertheimer 2010; Wertheimer 2003). The first condition demands that an agent's consent is informed, which requires them to have knowledge of the relevant facts about the act they are consenting to, and to comprehend what they are consenting to on the basis of these facts. Secondly, consent must be given voluntarily, in which case there is no unduly influence, such as coercion, from another agent, which is causally linked to the act of consenting. Thirdly, the consenting agent must be decisionally-capacitated, which requires them to have a grasp of the value and consequences of the act, on the basis of which they assess the risks, harms and benefits involved in this act.

Once an agent who intends to consent to a certain act meets these conditions, they must express their consent to make it recognizable to others. This requires something more than a mental state or attitude of consent enjoyed by the consenting agent, since this would pave the way for mistaken postulations of consent. What is required for others to justifiably believe that an agent has consented is that this agent has *performed* their consent. They can do so either verbally or non-verbally. A non-verbal performance of consent can occur with the performance of certain actions that do not involve language, such as nodding or initiating a certain activity. An act of consent can be performed verbally either in writing, for instance in signing a contract, or in speech, in uttering "yes" or its equivalent (Kleinig 2010; Mappes 2002; Miller and

Wertheimer 2010; Wertheimer 2003).[7] However, the criteria for the successful per-formance of consent are extremely complex, varying with the relationships between the agents involved and the act to which an agent is consenting, and involving many other complicating factors on which there is ongoing academic disagreement.

*2.2 Sexual Non-consent*

There are at least two categories of sexual non-consent. The first category con-cerns cases in which the agents involved in sexual activity are capable of expressing valid consent, but one or more has withheld their consent. These are paradigmatic cases of rape. The focus of this essay, however, is on a second category, which in-volves cases in which at least one agent's consent to sex is lacking because they are, either temporarily or permanently, cognitively incapacitated to give valid consent at all. Agents who lack or are incapable of exercising the cognitive capacities to meet the conditions for valid consent include non-human animals, children, and indi-viduals who are severely cognitively disabled, unconscious or heavily intoxicated (henceforth, 'cognitively incapacitated humans and non-human animals'). The ca-pacities these agents lack or fail to exercise may be, amongst other things, the capac-ity for rational deliberation, moral understanding, future-orientedness, or conscious experience.

The group of agents whose involvement in sexual activity is non-consensual because they are cognitively incapacitated to give valid consent includes sexbots as well. As became clear in the previous section, sexbots also lack the cognitive capaci-ties necessary for valid consent, such as the capacities for moral understanding and conscious experience. Since valid consent is a necessary condition for permissible sexual activity, it seems to follow from this that the valid consent condition does not only rule out permissible sex with humans and non-human animals, but also with sexbots.

This implication will strike many as counterintuitive. One initial response could be that the implication can be avoided simply by programming sexbots to perform actions that indicate consent, or even refusal of consent. For example, they could be

---

7.    One helpful way to conceive of verbal performative consent is to understand it in terms of J. L. Austin's speech act theory (Austin 1973). A speech act is an utterance that has not only semantic meaning ('locution') or causal effects ('perlocution'), but is itself an act, performed by the agent in uttering a certain phrase ('illocution'). Thus, on this understanding, an agent can perform the act of consenting in uttering certain words. For such an interpretation of performative consent, see for example Marta (1996) and Cowart (2004).

coded to utter phrases that under the right conditions would express sexual consent, or to sometimes struggle as an indication of withheld consent. However, such a response fails. Sexbots lack the cognitive capacities to satisfy the conditions for valid consent even when they are programmed to perform actions that would under the right circumstances constitute acts of consent. Performances of consent can be genuine acts of consent only if the cognitive conditions for valid consent have been met in the first place. Since sexbots are cognitively incapacitated to meet these conditions, their performances are failed acts of consent, devoid of any normative force.

Merely programming sexbots in a way that causes them to display a void performance of consent or non-consent is thus unsuccessful in showing that the valid consent condition does not render sexual relationships between humans and sexbots impermissible. Rather, if the conclusion that sex with robots is impermissible because it is non-consensual is to be avoided, without denying that valid consent is necessary for permissible sex with humans or animals, it must be demonstrated that there is a morally significant difference explaining why consent is a necessary condition for permissible sex if it involves humans or animals, but not if it involves robots. In the next section, I will discuss and reject three possible ways to do this.

## 3. NON-CONSENSUAL SEX: ROBOTS VERSUS HUMANS/ANIMALS

### 3.1 Artificiality

One way to argue that there is such a morally significant difference is to stress that valid consent is required only from members of the human species, in virtue of their humanity. Yet to ascribe such moral weight to the humanity of the parties involved is not only in itself questionable, on suspicion of unfounded speciesism, but also implies that all sexual activity with non-human animals is morally permissible. To avoid this implication, the argument can be modified so that the perceived moral difference between non-consensual sex with humans and non-human animals on the one hand, and non-consensual sex with robots on the other hand, is explained by the fact that the former are organic and the latter artificial. However, if the distinction between organic and artificial were morally relevant in this way, the incorporation of artificial body parts into a human body would diminish the moral weight of

this human being's interests. This would mean that the more artificial body parts a human being has in their body, the less their consent matters. This is clearly absurd.

Finally, if the appeal to the distinction between organic and artificial bodies is supposed to indicate a moral difference between living and lifeless entities, life as such is deemed morally relevant for consent. Such a view would be highly implausible, given that the demand for consent does not apply, nor should it, to *all* living entities, in virtue of being alive. However, if the claim is weakened as to assert that lifeless entities are essentially devoid of value, whereas *some but not all* living entities are valuable in a way that generates a moral incentive to seek their consent, the morally significant difference is no longer that between living and lifeless entities.

### 3.2 Harmlessness

Another possible line of argument could be that the morally significant difference between non-consensual sex with a human or non-human animal and non-consensual sex with a robot is that it harms the former but not the latter. I will understand the notion of harm in the comparative sense, since this understanding is most likely to capture the complex way in which non-consensual sex can cause harm to non-consenting subjects, and is widely employed in ethical discourse. On the comparative understanding of harm, an act X harms an agent S if S is worse off as a consequence of X. In other words, X harms S if S would have been better off if X had not been performed.

Non-consensual sex can cause harm to non-consenting subjects in many ways; physically, and particularly psychologically. For example, non-consensual sex is likely to involve the experience of pain in cases in which sex is forced on an individual who has refused to give consent, although this is not true for every instance of non-consensual sex. Another way in which non-consensual sex is often harmful to the individual whose consent is lacking is is in objectifying her as a mere means to sexual gratification, rather than as an end in themselves (Archard 2007; Baber 2002; Mappes 2002).

There is, of course, much more to say about the way in which non-consensual sex is harmful to those who undergo it. Nonetheless, it is already clear at this point that involvement in non-consensual sexual relationships does not cause harm to sexbots, since sexbots lack the capacity to experience pain and, as objects, are not harmfully objectified. More generally, it seems that it is impossible for sexbots to be

harmed at all. For things can go better or worse for entities only if they have qualitative experiences and enjoy some degree of consciousness. Since sexbots are non-sentient entities that lack the capacity for conscious experience, they cannot be harmed.

Still, however, it does not follow from the mere fact that an act of non-consensual sex is harmless that it is permissible. To see this, consider the following case:

*Case I*

*A is a child who is unconscious at T1. B has non-consensual sex with A at T1, which B considers to be an act of love. There are no third parties involved or present who could report about the act, and there will be no other evidence of the sexual act, such as bodily traces. A will therefore never know that B had non-consensual sex with them at T1.*

In this case, A will experience neither physical nor psychological effects of B's act of non-consensual sex with them at T1. Moreover, the act of non-consensual sex does not harmfully objectify A as a mere means to sexual gratification, since B considers it to be an act of love by which they value A as an end in themselves. Consequently, there is no reason to suppose that A is made worse off by this act. If A is not made worse off by the sexual act, the act does not harm them. Still, B's act of non-consensual sex with A at T1 is seriously wrong. This shows that the mere fact that an act of non-consensual sex is harmless is insufficient to render it permissible.

*3.3 Cognitive Deficiency*

One could also argue that non-consensual sex with robots is permissible but non-consensual sex with humans or non-human animals is not, for the reason that, compared to humans and non-human animals, robots are too far down the scale of cognitive capacitation. In other words, the cognitive gap between robots on the one hand, and humans and non-human animals on the other hand, is large enough to cause a discrepancy in the permissibility of non-consensual sex. The explanation of this discrepancy would then be grounded in the fact that humans and non-human animals are cognitively more developed than robots, and hence approach the threshold for valid consent more closely than robots.

However, although cognitive capacities may come in degrees, valid consent is

binary rather than scalar: one has either given valid consent, or one has not, depending on whether or not the conditions for valid consent have been satisfied. To illustrate this with respect to valid sexual consent, consider the following case:

*Case II*

*A has sex with B and C has sex with D. Both B and D are severely cognitively disabled and incapable of giving valid consent to these sexual acts. B is more heavily cognitively disabled than D, because B has more developed cognitive capacities than D.*

In this case, both sexual acts are non-consensual, since both B and D are cognitively capacitated to a degree that is insufficient to ground valid consent. However, on the account under consideration, the comparative permissibility of these acts would be affected by the fact that B finds themselves higher up the scale of cognitive capacitation than does D, since B more closely approaches the threshold for valid consent. This is an implausible view, since both acts of non-consensual sex seem equally wrong. This case thus illustrates that valid consent is binary, and that the degree to which an individual is cognitively capacitated does not matter for the permissibility of sexual activity when this degree is insufficient to ground valid consent. Hence, there is no significant moral difference with respect to the permissibility of non-consensual sex between having the requisite capacities for valid consent to a certain insufficient degree and lacking these capacities completely.

## 4. MORAL STATUS

Still, one could argue that the morally significant difference between robots one the one hand, and cognitively incapacitated non-human animals and humans on the other hand, is that the latter possess the requisite cognitive capacities to an insufficient degree, whereas the former are *devoid* of these capacities altogether. In the case of robots, the cognitive capacities necessary for valid consent are not just rudimentary, distorted, or underdeveloped, as is true of cognitively incapacitated humans and non-human animals, but absent in the past, present and future. However, such an account does not function as an explanation of the discrepancy in the permissibility of non-consensual sex between cases in which the non-consenting subject is a

robot and cases in which this subject is a human or non-human animal. The mere fact that robots are not on the scale of cognitive capacitation for consent at all, whereas humans and non-human animals are on this scale on a level too far down to meet the conditions for valid consent, does not single out a fact that is in itself morally significant.

Nonetheless, this fact does become morally significant if it is taken to be indicative of moral status. On certain established accounts of moral status, sentience is both a necessary and sufficient condition for moral status (Kagan 2016; McMahan 2002; Singer 1993). Sentience is usually defined as the capacity to have qualitative experiences, most importantly of pleasure and pain. Some philosophers argue that sapience is an additional ground for moral status (Kagan 2016; McMahan 2002). The notion of sapience refers to an entity's capacity to enjoy a degree of psychological continuity, which originates from certain sophisticated cognitive phenomena, such as self-awareness, future-orientedness and moral understanding (Kagan 2016; McMahan 2002).

Since sexbots lack qualitative experiences, self-awareness, future-orientedness, and moral understanding, amongst other things, they are both non-sentient and non-sapient. Sexbots are therefore devoid of moral status. To say that sexbots lack moral status is to say that they do not matter morally for their own sake. If sexbots do not matter morally for their own sake, they are not the kind of entities that humans require consent from in order to do things to them, such as having sex with them. Because humans do not require consent from sexbots, as entities that lack moral status, sexual relationships between humans and sexbots are not impermissible in virtue of being non-consensual.

By contrast, insufficiently cognitively capacitated humans and non-human animals do have moral status, since they are sentient and, to varying degrees, also sapient. Because they have moral status, they are entities we require consent from in order to do things to them, such as to engage in permissible sexual activity with them. Hence, when insufficiently cognitively capacitated humans and non-human animals are, as a matter of fact, incapable of consenting, others must refrain from having sex with them.

The morally significant difference between non-consensual sex with humans and non-human animals on the one hand, and robots on the other hand, thus seems to be that we require consent from entities that have moral status, such as humans and non-human animals, but not from entities that lack moral status. Since robots

are devoid of moral status, humans do not require consent from sexbots in order to have permissible sex with them. This explains why valid consent is a necessary condition for permissible sex in the case of humans and non-human animals, but not in the case of robots.

## 5. CONCLUSION

My aim in this essay has been to provide a plausible argument to show that there is a morally significant difference between non-consensual sex with robots and non-consensual sex with humans and non-human animals, which substantiates and justifies the intuition that the latter is impermissible but the former is not. I argued that the relevant moral difference is that robots are devoid of moral status, whereas humans and non-human animals are not.

Admittedly, this solution holds only insofar as sexbots are non-sentient and non-sapient. The conclusion of this paper is therefore conditional on scientific and technological developments in computer science, artificial intelligence and robotics, which determine the level of cognitive sophistication enjoyed by robots. Indeed, would sexbots acquire some degree of sentience or sapience in the more distant future, and start to matter morally for their own sake, their design, sale and use for the purpose of providing sex for humans might become seriously wrong.

Finally, it is worth emphasizing that even if it is not impermissible to have sex with near future sexbots for the reason that it is non-consensual, it might well be impermissible for other reasons. As indicated in section 2, there are legitimate worries about harmful effects that sexual relationships with sexbots might have on others, particularly on women, as well as the intrinsic moral wrongness of certain behaviour that the availability of sexbots might facilitate. However, more philosophical reflection is needed to settle these issues, and to establish under what conditions, if at all, it would be permissible for humans to have sex with sexbots.

## REFERENCES

Archard, D. "The Wrong of Rape". *The Philosophical Quarterly 57*, no. 228 (2007): 374-393.

Austin, J. L. *How to Do Things with Words*. New York: OUP, 1973.

Baber, H. E. "How Bad is Rape?" *The Philosophy of Sex: Contemporary Readings*. Edited by A. Soble. Lanham/Oxford: Rowman & Littlefield, 2002.

Cahill, A.J. *Rethinking Rape*. Ithaca/London: Cornell University Press, 2001.

Cowart, M. "Understanding Acts of Consent: Using Speech Act Theory to Help Solve Moral Dilemma's and Legal Disputes". *Law and Philosophy* 23, no. 5 (2004): 495-525.

Cowling, M. et al. *Making Sense of Sexual Consent*. Routledge, 2017.

Danaher, J. "Robotic Rape and Robotic Child Abuse: Should they be Criminalized?". *Criminal Law and Philosophy* 11, no. 1 (2017): 71-95.

Devlin, K. "In Defence of Sex Machines: Why Trying to Ban Sex Robots is Wrong". September 17, 2015. https://theconversation.com/in-defence-of-sex-machines-why-trying-to-ban-sex-robots-is-wrong-47641 [Accessed 20 July 2017].

Giutu, S. "Sex Robots and the Roboticization of Consent". We Robot Conference, April 21-22, 2012. Retrieved from http://robots.law.miami.edu/wp-content/uploads/2012/01/Gutiu-Roboticization_of_Consent.pdf.

Levy, D. *Love + Sex with Robot*s. London: Duckworth, 2009.

Li, J., Ju, W. and Reeves, B. "Touching a Mechanical Body: Tactile Contact with Intimate Parts of a Humanoid Robot is Physiologically Arousing" (2016). Retrieved from http://www.wendyju.com/publications/li_mechanical.pdf.

Kagan, S. "What's Wrong with Speciesism?" *Journal of Applied Philosophy* 33, no. 1 (2016): 1-21.

Kleinig, J. "The Nature of Consent". *The Ethics of Consent*. Edited by F.G. Miller and A. Wertheimer. New York: OUP, 2010.

Mappes, T. "Sexual Morality and the Concept of Using Another Person". *The Philosophy of Sex: Contemporary Readings*. Edited by A. Soble. Lanham/Oxford: Rowman & Littlefield, 2002.

Marta, J. "A Linguistic Model of Informed Consent". *The Journal of Medicine and Philosophy* 21, no. 1 (1996): 41-60.

McMahan, J. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: OUP, 2002.

Miller, F.G., Wertheimer, A. et al. *The Ethics of Consent: Theory and Practice*. Edited by F.G. Miller and A. Wertheimer. New York: OUP, 2010.

Nesta FutureFest Survey. April 27, 2016. Retrieved from http://www.comresglobal.com/polls/nesta-futurefest-survey-2/ [Accessed 23 July 2017].

Richardson, K. "The Asymmetrical 'Relationship': Parallels Between Prostitutions and the Development of Sex Robots". *Computers and Society* 45, no. 3 (2015): 290-293.

———"Sex Robot Matters: Slavery, the Prostituted, and the Rights of Machines". *IEEE Technology and Society Magazine* 35, no. 2 (2016): 46-53.

Rutkin, A. "Could sex robots and virtual reality treat paedophilia?". *New Scientist*, 2 August 2016. https://www.newscientist.com/article/2099607-could-sex-robots-and-virtual-reality-treat-paedophilia/ [Accessed 20 July 2017].

Soble, A. et al. *The Philosophy of Sex: Contemporary Readings.* Edited by A. Soble. Lanham/Oxford: Rowman & Littlefield, 2002.

Scheutz, M. and Arnold, T. "Are We Ready for Sex Robots?". *The Eleventh ACM/IEEE International Conference On Human Robot Interaction* (2016): 351-358. Retrieved from https://hrilab.tufts.edu/publications/scheutzarnold16hri.pdf.

Sharkey, N. et al. *Our Sexual Future with Robots: A Foundation for Responsible Robotics Consultation Report.* May 2017. http://responsiblerobotics.org/wp-content/uploads/2017/07/FRR-Consultation-Report-Our-Sexual-Future-with-robots_Final.pdf.

Singer, P. *Practical Ethics*. Cambridge: CUP, 1993.

Sparrow, R. "Robots, Rape and Representation". *International Journal of Robotics* (online first, 2017).

Sullins, J. "Robots, Love, and Sex Machines: The Ethics of Building a Love Machine". *IEEE Transactions of On Affective Computing* 3, no. 4 (2012): 398-409.

Szczuka, J. and Krämer, N. "Not Only the Lonely – How Men Explicitly and Implicitly Evaluate the Attractiveness of Sex Robots in Comparison to the Attractiveness of Women, and Personal Characteristics Influencing this Evaluation". *Multimodal Technologies and Interaction* 1, no. 3 (2017): 1-18.

Wertheimer, A. *Consent to Sexual Relations*. Cambridge: CUP, 2003.

# Prostitution: You Can't Have Your Cake and Sell It

SIMON-PIERRE CHEVARIE-COSSETTE

*University of Oxford*

2017 OXFORD UEHIRO PRIZE IN PRACTICAL ETHICS

HONORABLE MENTION

## ABSTRACT

I offer an unorthodox argument for the thesis that prostitution is not just a normal job. It has the advantage of being compatible with the claim that humans should have full authority over their sexual life. In fact, it is ultimately the emphasis on this authority that leads the thesis that prostitution is a normal job to collapse. Here is the argument: merchants cannot (both legally and morally) discriminate whom they transact with on the basis of factors like the ethnicity or the religion of their client; but if prostitutes are 'sex merchants', then they cannot (both legally and morally) discriminate whom they have sex with on the basis of these factors. Yet everyone should have the full discretionary power to refuse to have sex under any circumstances.

## 1. INTRODUCTION

You have made it thus far: the wedding preparation is almost over. You enter your local bakery, cheekily anticipating the moment when you'll order a wedding

cake for 'John & John'. But to your dismay, the baker turns you down because your marriage goes against his 'Christian beliefs'.

This is a true story and it is a recurrent one. In 2013, Administrative Law Judge Robert N. Spencer found the owner of Masterpiece Cakeshop guilty of discrimination on the basis of sexual orientation.[1] The decision was then maintained by the Colorado Civil Rights Commission[2] and again by the Court of Appeal[3]. The Supreme Court of Colorado refused to reconsider the case.[4] Analogous situations have occurred in Texas and in Northern Ireland.

Discrimination occurs, unfortunately, much more frequently than homosexual weddings. Thus it has been the object of laws around the globe. For instance, the article that Spencer invokes reads:

> *It is a discriminatory practice and unlawful for a person, directly or indirectly, to refuse, withhold from, or deny to an individual or a group, because of disability, race, creed, colour, sex, sexual orientation, marital status, national origin, or ancestry, the full and equal enjoyment of the goods, services [...]. (Colorado Anti-Discrimination Act: 24-34-601, (2) (a))*

In 24-34-601, discriminatory practices are defined in the context of trade within a business establishment or a public institution, but other similar laws[5] apply to all trade indistinctly.

Sometimes discrimination is allowed. For instance, on safety grounds, a dwarf may be denied access to a rollercoaster and a leper may be refused a chiropractor's massage. With this qualification in mind, we can introduce the following principle:

> *Non-Discrimination*: A tradesman has the enforceable obligation not to refuse goods or services to clients on the basis of their disability, religion, sexual orientation, etc., unless this trait makes it dangerous for them to receive the service or possess the good.

---

1.    See Craig & Mullins vs Masterpiece Cakeshop Inc. & Phillips; CR 2013-0008, December 2013.

2.    See the Colorado Civil Rights Commission's decision, CR2013-008, May 2014.

3.    See Court of Appeals No. 14CA1351, August 2015.

4.    See Supreme Court Case No. 2015SC738, April 2016.

5.    See, for instance the Equality Act 2010 of British Law or ch. I, 1., 10-12. of Québec's Charter of Human Rights and Freedoms.

Thereby the State should intervene, were it to recognise relevant discrimination. This seems plausible: discrimination can directly cause severe harm to individuals—harm they could not resist on their own. Allowing discrimination might also indirectly harm other individuals by giving rise to a toxic social environment.

Surprisingly, the cake story teaches us a lesson about prostitution: if the Court's ruling is correct, then *prostitution is not a job like any other*.[6] This conclusion follows from the inclusion of sex in the list of 'goods and services' mentioned in *Non-Discrimination* together with some equally plausible principles.

§2 describes why some have taken prostitution to be a normal job. §3 argues against this from *Non-Discrimination* and addresses some objections. §4 concludes.

## 2. THE DEBATE OVER PROSTITUTION

Why should prostitution be recognised as a job and prostitutes as sex workers? Transactions should be legal if (and only if[7]):

1. The transaction is voluntary;

2. The transaction does not directly harm people who are not parties to it[8];

3. The transaction does not cause an unacceptable kind or degree of harm to any of the parties of the transaction.

Call *liberalism about prostitution* the view that sex transactions involved in prostitution satisfy or could in principle satisfy these three criteria. Liberals about prostitution argue as follows. Firstly, people often get involved voluntarily in prostitution. When the sexual transactions are involuntary, this is 'rape' or 'human trafficking', not 'prostitution'. Secondly, no one else is involved in the sex transaction: it is a service delivered by an individual to another, in a private environment. Thirdly, the harm of prostitution to the prostitute is comparable to the harm of many other stressful jobs

---

6. After the completion of this paper, the author discovered that a comparable approach was taken in a blogpost to defend a similar claim (Watson 2014).

7. Although it seems reasonable to add 'only if', it is not strictly necessary for my argument.

8. Differently put, those who do not consent should not be directly harmed (cf. Saunders 2016, 1111). Of course, what counts as direct harm needs to be strictly limited: for instance, the man who engages in a voluntary sexual transaction with his wife's friend may harm his wife in one sense.

(Nussbaum 1999, 288–97). Voluntary activities that fail to satisfy the third criteria are very rare (e.g. cannibalistic transactions[9]) and do not include prostitution.

The liberal about prostitution then normally claims something along the following line. If the transactions that make up a job meet conditions (1), (2), and (3), then this is sufficient for a job being a normal job in the sense that it should be legalised and regulated according to the general rules of trading goods and services[10]. Therefore, liberals about prostitution endorse:

> *Normal Job*: Prostitution is a job like any other, i.e. it should be legalised and regulated, in general, in accordance with the rules of trading goods and services.

It is important to note that *Non-Discrimination* is not only one of these rules; it is a crucial rule of trading goods and services. By contrast, there may be non-crucial rules of trading goods and services which need not apply for some normal jobs.

Non-liberals about prostitution normally object to *Normal Job* on the grounds that prostitution does not satisfy (2) and (3). They firstly argue that the transaction harms people outside of the sexual transaction because it transforms the nature of non-market sexual relationships (E. Anderson 1993, 154–55), for instance by representing women as sexual servants of men (Satz 1995, 78). Non-liberals secondly argue that selling sex is extremely harmful to the prostitute because it expresses a lower social status and a loss of identity (Pateman 1988, 207) or because it limits the prostitute's autonomy (S. A. Anderson 2002, 386).

The disagreement over (2) turns principally on which harms should be recognised as such by the State. For instance, if I hurt my wife by kissing another woman, it is not the State's business. Liberals think that the same goes for the sexist symbols conveyed by prostitution. For our purpose, I will leave this question aside.

The disagreement over (3) turns on whether sexuality is special to human integrity. Selling sex is degrading, claims the non-liberal. The liberal disagrees: it is up to anyone to determine the meaning of sex in their life. State intervention only worsens the lives of prostitutes. More generally, the liberal endorses:

---

9.    See the case of Detlev Günzel (http://www.bbc.com/news/world-europe-32146031).

10.    When it is explicitly laid out, this premise may sound unacceptable to some liberals about prostitution. This does not however prevent them from often making the claim that prostitution is just a normal job, which is the main thesis under evaluation.

*Liberal Sexuality*: The State is permitted to prevent, or punish one of the parties to, a sexual transaction if and only if this transaction is or was involuntary.

This principle enshrines the absolute discretionary power that each has over their sex life.

## 3. DISCRIMINATION OF CLIENTS: SEX AND CAKE

Liberalism about prostitution gives us an appealing picture thus far: because it involves voluntary transactions which do not directly harm others and which are not unacceptably harmful, prostitution is a normal job. This claim is prima facie coherent with our intuition that the State has no business in regulating voluntary sexual transactions. Nevertheless, this picture is incompatible with the lessons of the cake story. Consider again:

*Normal Job*: Prostitution is a job like any other, i.e. it should be legalised and regulated, in general, in accordance with the rules of trading goods and services.

*Non-Discrimination*: A tradesman has the enforceable obligation not to refuse goods or services to clients on the basis of their disability, religion, sexual orientation, etc., unless this trait makes it dangerous for them to receive the service or possess the good.

*Liberal Sexuality*: The State is permitted to prevent, or punish one of the parties to, a sexual transaction if and only if this transaction is or was involuntary.

These three principles are jointly untenable. *Normal Job* and *Non-Discrimination* together entail that prostitutes have an enforceable obligation not to refuse service to clients on the basis of their disability, religion, etc. Now assume that a prostitute refused a sexual service to a client on such a basis. The prostitute might find it sickening to sleep with a married man; she might detest having sex with a fundamentalist; whatnot. According to *Non-Discrimination*, the state may intervene (to require that

she stop discriminating and compensate the client). But *Liberal Sexuality* tells us that the State may *not* intervene because there has *not* been an involuntary sexual transaction. The prostitute is free to refrain from having sex *regardless of her reasons*. To avoid this contradiction, I suggest that we abandon *Normal Job*.

A natural response to this argument is, for those who refuse to amend *Normal Job*, to insist that the State intervention in such a case is not strictly speaking about sex. The State does not force the prostitute not to discriminate, but rather not to discriminate *if she wants to keep her job*. The object of the enforced obligation is thus not sexual. It reads: '*If you want to keep being a prostitute*, then ...'. Thereby, *Normal Job, Non-Discrimination*, and *Liberal Sexuality* are compatible.

This reply at best defers the refutation of *Normal Job*. In a word, if the three principles are true, it allows the State to create some exploitative situations, which is absurd. Consider two additional principles:

*Exploitation*: One's being forced either to have sex with someone whom one does not want to have sex with or to quit one's job is a case of an exploitative situation.[11]

And

*Role of State*: The State should never create exploitative situations.

*Role of State* is minimal: most liberals think that the State should not only refrain from creating but also fight (directly or indirectly) exploitative situations. *Exploitation* is also difficult to deny. It warrants our judgement that teachers and bosses should generally not have sex with their students or their employees. This is because there is a high risk for a person in authority of putting her subordinates in an exploitative situation or in a situation that is perceived as such.

Note that if a prostitute in the situation described in *Exploitation* decided to have sex nonetheless, without changing her mind about what she wanted, she would act intentionally, but not voluntarily. If I decide to hand in my wallet to a burglar because I do not want to risk my life, I do so intentionally but involuntarily (Hyman 2015, 87-91).

---

11.    I do not wish to rely on a specific account of exploitation here. Yet, that Exploitation turns out true is seemingly an important desideratum of any satisfactory account of exploitation.

*Exploitation, Role of State, Non-Discrimination, Liberal Sexuality,* and *Normal Job* are jointly contradictory. Here is how. A prostitute who would discriminate on the basis of her client's sex, religion, etc. would infringe her enforceable obligations (by *Non-Discrimination and Normal Job*). At this point, the State cannot directly intervene to force the prostitute to have sex (by *Liberal Sexuality*). Rather, it may send the message to the prostitute: 'have sex against your will or abandon your job'. The State has thus created an exploitative situation (by *Exploitation*). But the State should never do so (by *Role of State*). To avoid this contradiction, I suggest that we abandon *Normal Job*, for it is the weakest link of the reasoning.

Importantly, my argument does *not* assume that there are voluntary sexual interactions that the State should recognise as degrading. The argument simply maintains that the State should intervene when the interaction is involuntary; it says nothing about undesired, unemotional, non-committal sex. As such, it is compatible with *Liberal Sexuality*. (Note that the argument does not assume the opposite either, i.e. that there are no voluntary sexual interactions that the State should recognise as degrading. The argument remains neutral on this general topic.)

This becomes clearer once we distinguish the claim that voluntary sex is a special good (which the liberal denies) and the claim that involuntary sex is a special wrong (which everyone should admit). You do not need to think that sex is particularly transcendent to agree that being forced to choose between unemployment and sex with an individual that you do not want to transact with is exploitative. Some might point out that *Exploitation* should also be applied to similarly harmful tasks. I fully agree: if prostitution were just like some normal jobs, it would mean that these jobs were exploitative.

## 4. CONCLUSION

I am aware that my argument might look contrived to people who are concerned with the serious injustices of the sex trade. Clients who are discriminated against are certainly not the actual victims of this system. But this was not the point: I rather wanted to show the tension amongst liberal tenets on prostitution and discrimination. If we think that one should retain the absolute power to refrain from having sex and if we agree with the Court of Colorado's enforcement of civil rights, we should resist the idea that prostitutes are normal tradesmen.

If prostitution is not a normal job, what should the liberal say about sexual trans-

actions? On the one hand, the liberal may admit that sexual transactions should be directly or indirectly outlawed. It follows that either sex transactions do not satisfy (1), (2), and (3) (the alleged conditions for legalisation) or these criteria are insufficient. On the other hand, the liberal may insist that sexual transactions be legalised. For instance, individuals may trade sex, but it does not make them normal traders. Prostitutes could be protected workers, but who do not have legal duties regarding discrimination. This would make it a very special job indeed.[12]

## REFERENCES

Anderson, Elizabeth. 1993. *Value in Ethics and Economics*. Cambridge, Mass ; London: Harvard University Press.

Anderson, Scott A. 2002. "Prostitution and Sexual Autonomy: Making Sense of the Prohibition of Prostitution." *Ethics* 112 (4): 748–80.

Assemblée nationale du Québec. 1975. *Charte Des Droits et Libertés de La Personne*.

BBC News. 2015. 'Germany "Cannibal" Trial: Former Policeman Is Sentenced', April 1. http://www.bbc.com/news/world-europe-32146031 [Accessed November 2017].

Hyman, John. 2015. *Action, Knowledge, and Will*. Oxford: Oxford University Press.

Nussbaum, Martha Craven. 1999. *Sex & Social Justice*. Oxford: Oxford University Press.

Pateman, Carole. 1988. *The Sexual Contract*. Stanford, Calif.: Stanford University Press.

Satz, Debra. 1995. 'Markets in Women's Sexual Labor'. *Ethics* 106 (1): 63–85.

Saunders, Ben. 2016. "Reformulating Mill's Harm Principle." *Mind* 125 (500): 1005–32.

The Parliament of the United Kindgom. 2010. *Equality Act* 2010.

The State of Colorado. n.d. *Colorado Anti-Discrimination Act*.

Watson, Lori. 2014. 'Why Sex Work Isn't Work'. http://logosjournal.com/2014/watson/ [Accessed November 2017].

---

12.    Further work needs to be done on comparing prostitution with other 'special jobs'. For instance, a Catholic priest can refuse to conduct a funeral service for a Protestant without breaking the law, and a mohel can refuse to circumcise a non-Jew even though they may charge a fee for their service. But perhaps such exception is justified by the fact that religious organisations are recognised as non-profit organisations. As a matter of fact, discrimination is sometimes allowed – or rather, some exclusions are sometimes not considered discriminatory. For instance, Article 20 of the Quebec Charter of Human Rights and Freedom reads, 'A distinction, exclusion or preference based on the aptitudes or qualifications required for an employment, or justified by the charitable, philanthropic, religious, political or educational nature of a non-profit institution or of an institution devoted exclusively to the well-being of an ethnic group, is deemed non-discriminatory'. Clearly, Article 20 is not relevant to prostitution. Should there be additional restrictions analogue to those mentioned in Article 20 which would deal with prostitution? This may be the liberal's best avenue.

# The Ethics of Political Bots: Should We Allow Them For Personal Use?

JONAS HAEG

*University of Oxford*

2017 OXFORD UEHIRO PRIZE IN PRACTICAL ETHICS

RUNNER-UP FINALIST: GRADUATE CATEGORY

ABSTRACT

The technology to create and automate large numbers of fake social media users, or "social bots", is becoming increasingly more accessible to private individuals. This paper explores one potential use of the technology, namely the creation of "political bots": social bots aimed at influencing the political opinions of others. Despite initial worries about licensing the use of such bots by private individuals, this paper provides an, albeit limited, argument in favour of this. The argument begins by providing a prima facie case in favour of these political bots and proceeds by attempting to answer a series of potential objections. These objections are based on (1) the dangerous effectiveness of the technology; the (2) corruptive, (3) deceitful and (4) manipulating nature of political bots; (5) the worry that the technology will lead to chaos and be detrimental to trust online; and (6) practical issues involved in ensuring acceptable use of the technology. In all cases I will argue that the objections are overestimated, and that a closer look at the use of political bots helps us realise that using them is simply a new way of speaking up in modern society.

## 1. INTRODUCTION

A "bot", in the broadest sense, is any computer program built to perform automated tasks. While bots can be embodied in, and control, actual robots, most of them are not and work merely as a software or application on a computer or computing device. The application in charge of making your digital clock update the time on the screen every minute would constitute a bot in this sense, as it merely involves a repetitive, automated task (i.e. changing numbers on the screen). A subset of bots called "internet bots" are automated programs which operate *online* (for instance, gathering information about visitors on websites); and a further subset of these, called "social bots", operate on social media platform such as Twitter. These bots are programmed to run on, or control, their own social media profiles (identical to the ones you and I would have) and their tasks are of the social media type: posting messages, "liking" other posts, "following" individuals, and so on. While some of them are transparently bot-like (e.g. profiles programmed merely to share statistics about each vote in the U.S. Congress), others are designed to *mimic* human behaviour and to be perceived as genuine. Hence, the content of their messages, the patterns of posting, etc. are close to ours: emotional and personal, containing opinionated beliefs, and so on.

Within this latter category of social media bots, we find the bots that are the main focus of the present paper. The term "political bots" refer to human-looking social media bots with a clear *political aim*: to influence political discussions and opinions. For simplicity, I will refer to these bots as "polibots". The use of these bots have increased in recent years. According to some estimates, polibots produced almost 20% of all election-related tweets on Twitter a few weeks prior to the election day of the last U.S. presidential election.[1] Researchers and others warn us that the use of polibots will only increase around the world.[2] As more and more people turn to social media to discuss politics, and use social media as an information source, these platforms are increasingly being dominated by hordes of bots that try to influence opinions. This evolution has attracted academic and political interest from many[3],

1.   Bessi and Ferrara (2016).
2.   See, e.g., Miller (2017); Guilbeault and Woolley (2016).
3.   See, e.g., Oxford's "Computational Propaganda Project": http://comprop.oii.ox.ac.uk/

but few philosophers have engaged with the topic. With this paper, therefore, I hope to start a philosophical discussion about this topic by investigating the ethics of using polibots.

The current academic and popular literature on polibots is highly sceptical of them—and rightly so.[4] Yet this is largely in part due to their isolated focus on only one sort of polibot, or rather, one sort of use of polibots. I distinguish, roughly, between three types of polibot uses. These are the *Malicious Polibots*, the *Personal Polibots*, and the *Saintly Polibots*. Current writings on the topic focus almost exclusively on the first kind. I call them Malicious because they are the sorts of bots devoted to obviously malicious aims such as sharing hate-speech, false information, suppressing certain voices, etc., often with the aim or intention of maliciously steering a certain election. On the other side are the *Saintly* kinds, which are devoted merely to sharing important, true and relevant political information, with no allegiance to particular issues or sides.[5] In one sense, I think the first kind is obviously objectionable and the last kind obviously good. However, this paper is concerned with the ethics of *Personal Polibots*. In short, this refers the use of polibot technology by individual, political agents as an extension of their political voices. This paper provides an ethical examination of such use of the technology. The guiding question throughout the paper, therefore, is whether we should allow private persons to use the polibot technology to enhance their political voices online, or whether even this sort of use of polibots is objectionable.

The first section elucidates further the nature of the polibots we are focusing on here, and provide a *prima facie* case for their permissibility. After that, I present a series of potential arguments against them. Some of these attempt to target polibot use as something inherently problematic. Others focus on the bigger picture and dangerous consequences. In all cases, however, I rebut the arguments with some concessions. The overall conclusion is that the case against Personal Polibots have shortcomings and that polibots should be seen as largely permissible and indeed potentially valuable.

---

4.    See, e.g., Woolley (2016); Newman and O'Gorman (2017), Andrews (2017), Lee (2017), Alfonso (2012), Finley (2015) and Hern (2017) for examples and sceptical voices.

5.    Another term for these are "transparency bots".

## 2. THE NATURE OF POLIBOTS

In the literature, "political bot" is used to refer to a wide range of different technologies and uses of those technologies. Some bots are programmed solely to "repost" and "like" certain other people's posts; and others merely devoted to "follow" certain figures to boost perceived popularity.[6] Others again are programmed to share their own content, and these seem to me to be the ones most clearly programmed to mimic genuine humans. They are relatively easy to set up. One needs to pre-write the content that is to be shared, and then decide how frequently the bot is supposed to post these tweets. Conceivably, they could be programmed to post every few seconds, but most likely people will want them to look more human and therefore lower the frequency of posts and also program in "off-time" to make it look like the bot is sleeping, working, etc.

As said in the introduction, even amongst the bots devoted to share their own content, there are large differences in what sorts of content and aims they can be programmed for. The *Malicious Polibots* are so called because they are obviously problematic. These bots, however, can be argued against on grounds independent of the polibot technology itself. After all, the explicit content and aim of these bots are types of content and aims we find morally objectionable in general. For instance, the sharing of hate-speech, fake news and information, trying to suppress certain voices, and so on. It is easy to see that these are morally problematic independently of the technology. On the other side are the *Saintly Polibots*. And likewise, it is easy to see why we should be in favour of their existence on grounds independent of the technology. After all, they are programmed to share non-biased and politically relevant information, which are often not shared widely in mainstream news sources.

However, my interest here lies in the ethics of *polibots* in themselves. The problems with *Malicious Polibots* and the good features of *Saintly Polibots* aren't problems or features connected to the *polibot* technology *per se*, but rather the specific content or aims they have. To delve deeper into the specific issue concerning the use *polibots* itself, I focus on what I call *Personal Polibots*. These seem to be to be neither obviously objectionable nor obviously morally good. I am assuming that they will not share obviously problematic content such a hate-speech and fake information, and that the aim isn't malicious in the sense of being intended to fool people to sway

---

6.    See, e.g, McKelvey and Dubois (2017) and Wooley and Guilbeault (2017) for more on different kinds of polibots.

their opinion, or try and scare away certain voices online. Rather, the polibots are programmed to share, in a non-offensive manner, information and opinions supporting the political cause of the programmer themself. In a sense, I take them to be programmed extensions of how many people are already using their personal social media accounts to engage in political matters and gain political influence. They are used to share the political beliefs and messages of their programmer. At the same time, this use isn't assumed to be morally saintly either. For in contrast with the objective, transparency-type polibots, normal people aren't neutral. They choose sides, posts information supporting their cause, and might refrain from posting information that would hurt their side. *Personal Polibots*, then, while required not to share hateful and false content, aren't required to be impartial. Focusing on this personal type of polibots will allow us to more clearly investigate the ethics of the technology itself, and whether we should permit its use or not.

Now, given the restrictions on what sorts of polibots we focus on, we can reasonably say that the sort of polibots of interest here are those which function as "communication tools". By this I mean tools that help us enhance our communication in different ways. After all, what is it that the polibots mentioned would add to the ordinary means that political agents have of sharing their political messages? They provide an enhanced scope of reach and potential for influence. Using one personal user profile online in real time gives you a certain level of reach and influence (depending on other factors as well, of course, such as amounts of followers, hashtags, and so on). But with the aid of polibots that can act much quicker and more constantly than a single person can, the scope and influence can be enhanced drastically. As one reaches more persons with one's message, it becomes more likely that one will reach someone who is prone to be influenced by that message as well. Now, this feature of enhancing communication is one that polibots share with many other political "communication tools" available today. Examples include hanging up political posters, writing op-ed pieces for newspapers, paying for political advertisements, and so on. Even using one personal social media account would be an example. As such, the polibots discussed here seems, on the face of it, to simply be a *new* communication tool available to people.

This provides, I think, a *prima facie* case for their permissibility. However, I suspect many people have some reservations also about this sort of use of polibots. Below, I outline and respond to some of the potential objections people might raise.

Although they highlight interesting ethical issues surrounding polibots, they do not, I argue, manage to show that we shouldn't permit them.

## 3. ARGUMENTS AGAINST THE PERSONAL USE OF POLIBOTS.

*3.1. Objection 1: Polibots Are Too Effective.*

Someone might argue that polibots are impermissible because they are "too effective" communication tools – that they will have too much influence. One might divide this worry into two separate ones. First, one might worry that there is such a thing as "too much influence" *simpliciter* and that polibots have this. Second, one might instead worry that polibots unfairly give *some* people *more* influence than others have.

The first worry might be put, somewhat metaphorically, like this. Everyone is entitled to a *political voice*, to engage in political matters with and try to convince one another. However, using polibots gives one, effectively, *several* political voices because they enhance one's message so much. But since we are only entitled to *one* voice, their use is impermissible. The first problem with this objection, however, is to explain why polibots go above some sacred threshold of political influence. Presumably, using polibots might be more effective than simply walking around the streets alone telling people your opinion. But there are many other sorts of communication tools that seem much more effective at influencing than this, which we nevertheless do allow. Examples might include writing op-ed pieces for the news, running a blog, or hanging up posters. If these are not above the aforementioned threshold, but still more effective than using no technology at all, it is hard to determine whether polibots are too effective. For showing that it is *more* effective than using no technology doesn't suffice to show that it is *too* effective.

Moreover, the objection is most likely false if we focus on the use of just *one* or a *few* polibots. For the use of only a few polibots will in most cases, I think, not be more influential than the use of many of the other permissible technologies mentioned above. Hence, this worry cannot be an intrinsic objection to polibots. Granted, however, the worry is more realistic when we consider the use of *many* polibots. Even so, showing that some use polibots can be more effective than other traditional communication tools does not entail that this is too effective. For "more than" doesn't entail "too much". Moreover, Woolley and Guilbeault (2017, 10) have already dis-

cussed this potential effect of polibots which they call the "megaphone" effect, i.e. the use of polibots to heavily amplify a political message. They also say that the worry is not really a polibot problem *per se*, but rather a problem with the *number* of polibots. Viewed in this light, however, we see that the worry isn't particular to polibots. For the megaphone worry arises for all sorts of communication tools. Compare paying for TV adverts, phone banking, and so on. All of these tools *can* become objectionably influencing when we start considering large numbers. For instance, we would object to someone buying up all air-time on TV to advertise for his candidate. The problem, however, isn't (necessarily) that we deem political TV adverts themselves impermissible tools, but rather that it is impermissible to exert so much influence. The same can go for polibots: they might not be impermissible in themselves, but their use can become impermissible if too many of them are used. At best, however, this objection would support some sort of regulation on polibots— something we're already familiar with from other communication technologies.

If we move to the second version of the worry, the same response probably hold. Recall, the worry was the polibots can give some people much more political influence than others. But again, this is a worry about the numbers and the scope of the availability of the technology. However, it also bears mentioning that polibot technology might actually have the opposite effect to some extent. People today are often worried that a few wealthy individuals gain disproportionately much political influence. But polibots are extremely easy to set up, and so the technology could potentially help democratize political influence somewhat by making it more available to those who cannot compete with rich persons on other arenas. Of course, though, money might buy one more effective bots than the average person can create themselves, and so the unfairness worry might continue as before. Again, though, as I said, this only highlights the potential need for some sort of regulation.

In sum, this objection does not speak forcefully against allowing the personal use of polibots. It points to a general problem with communication technologies, and it is not particular to polibots. Almost any sort of communication technology could conceivably be used so as to be "too effective". This doesn't show that any use of the technology is impermissible, and, moreover, it seems the correct response would be regulation. Lastly, it is worth noting that the *actual* effect and influence polibots have is an empirical question to which we currently lack clear answers. The current research, moreover, has tended to focus on coordinated polibot campaigns and not the

effects of private persons. It might turn out that it will be extremely hard for single individuals to have effects comparable to many of the more traditional means.

### 3.2. Objection 2: Polibots Corrupt Political Discourse.

Another potential objection is that polibots corrupt political discourse. Political discourse, the argument might go, is by definition limited to political *persons*. By activating polibots, one introduces non-persons into this discourse and therefore corrupts it. This objection is stronger than the above objection because it aims at showing that there is something intrinsically wrong about using even only *one* polibot.

The objection raises an interesting question about the ontology of polibots. In one sense, as I said earlier, a polibot appears as a mere tool or extension of the human controlling them. On the other hand, some might feel that they are more than this. Imagine, for instance, a social media platform of 100 genuine persons engaging in political discourse. Then, one of these people adds a polibot. It now seems as if there are 101 entities engaging on that arena, not all of whom are persons. There is some force to this intuitive picture, but let us attempt to dissect it: what is really the problem with that 101st entity?

Recall, polibots as here defined are very simple. A person needs to write and intend all of its posts beforehand. It is not really an intelligence in itself. Suppose, then, that I had the following two options: (1) stay at home for the weekend and post my regular political tweets, or (2) automate all of this in a script and take the weekend off. Returning to the picture above, it seems that choosing (2) would mean there are 100 entities on the platform, not all of which are persons. Yet, viewed in this light, it seems much less objectionable. I fail to see why the mere *automation* of one's political engagement itself is the problematic feature. The only difference between (1) and (2) seems to be that I am not posting my opinions *in real time*. But intuitively, this lack of real-time engagement seems insufficient to make it impermissible. This form of automatization doesn't, I hold, amount to introducing a distinct, non-person entity. Indeed, it seems that (2) has a lot in common with "auto-emails", i.e. emails typed beforehand and set to send at some later point. But that doesn't strike us as importantly different from writing emails in real time. In short, I cannot see why *automating* one's political engagement in this way corrupts the political discourse.

The opponent might, however, highlight two problems with the picture above. Polibots are different from (2) in two key ways: polibots usually come *in addition*

to one's personal account and have their *own identities*. It is not merely like a script running "your" online life (i.e. your personal profile); it adds seemingly disconnected users to that activity. I grant this. But, if one agrees with the above—that the corruption doesn't come from the mere automatization of political activity itself—it seems as if one must think that the corruption of political discourse comes from these two features: *additionality* and *anonymity*. This is implausible, however. It is already somewhat commonplace for person to have several social media accounts without necessarily disclosing this. Since these actions also involve additionality and anonymity, the current objection would equally apply to these people. However, the objection was that polibot activation are tantamount to introducing non-person entities online. But it seems outright implausible to say this in the case of someone activating users with additionality and anonymity (but which are controlled directly by the person). Clearly, this simply cannot be tantamount to introducing non-person entities; after all, they are themselves controlling these additional, anonymized users in real time.

Could it, however, be that it is only *anonymity* and *additionality + automatization* that results in corruption? Perhaps, though I find it unlikely that the mere, simple automatization could make this much of a metaphysical difference to the ontological status of those user profiles. In sum, therefore, I conclude that one cannot conclude that polibots are corrupting political discourse by introducing non-person entities.

### 3.3. Objection 3: Polibots Deceive.

Despite my pessimism above, I grant that there is another side to the worry. What matters, some might say, isn't the strict "ontology" of polibots but rather the perception that other people have of them. Using polibots without disclosing them as such, they might say, amounts to a form of *deception*. It is not the explicit content that is deceptive, but rather that people will be fooled by the activity of polibots and come to have false beliefs. In particular, they will believe that the bots are controlled by separate, genuine humans. As such, upon encountering my personal activity and my bot's activity, someone might wrongly infer that she has encountered *two* separate humans online, and in that way is deceived. Lynch (2016, 154-6), for instance, seems to think social bots in general are problematic for these sorts of reasons, claiming that bots are "massive deceit machines" that "operate on getting people to assume they are dealing with someone real who is sincere in their assertions".

However, I believe this worry is overestimated when we are dealing with *Personal*

*Polibots.* Suppose I have a polibot, B, which tweets the content "Donald Trump has been sued 135 times since taking office. Clearly he's not fit to be president!".[7] Deception is about causing false beliefs, but what belief is relevant here? It might be that *someone believes that Trump isn't fit to be president because of all the lawsuits*, but in a sense this belief is true. After all, *I*, the creator of B, believe so. In fact, it seems hard to spell out exactly what belief we are talking about. But let us grant that there is such a belief. Let us call it the belief that *B is a separate human who believes that ...* so that upon encountering both me and B, a person will make a wrong inference about how many persons believe the mentioned content. Has my activation of B constituted deception?

To deceive is traditionally defined as "to intentionally cause [someone] to have a false belief that is known or believed to be false" (Mahon 2015). On this account, however, it is questionable whether my polibot amounts to deception. For it is not clear that the false belief in question is part of the *intended* effect. It seems to me that the primary intention is not to get others to think that the polibot is a genuine human with certain opinions. Rather, the primary aim and goal is to get another person to access the explicit *information* (in the tweet) and become influenced by it. Of course, causing a (false) belief about who the bot is might be an effective means of doing so. After all, part of an effective way of getting someone to believe that *p* is to get them to believe that *someone sincerely believes that p*. But it seems that the agent behind *Personal Polibots* (as elucidated in Section 2) would be satisfied whether or not the other person saw through his bot, as long as she took information seriously and based her opinions partly on it. As such, they would pass the usual test for what is intended. Hence, I think the causing of the false belief will not be what is primarily intended. In many cases, it will be a mere unintended consequence.

Someone might respond that this misses a crucial fact: if people really didn't intend to deceive others about the identities of the polibots, they wouldn't use the polibots described here at all. For those bots are assumed to be anonymous. People who do not want to deceive would either mark their bots as bots or refrain from using the technology altogether. To respond to this point, we should look deeper at the reasons why people who intend to spread their message would nevertheless use these sorts of bots. One potential reason might be this. People want to spread their message as much as possible and so they seek technologies to help them do so. They also know that people are sceptical of bots (for instance, given all the negative press about bots, we are likely to be sceptical about who controls the bots, whether they are speak-

---

7.    See Hallemann (2017).

ing truthfully and so on). They therefore opt for anonymous polibots, because they aren't obviously identifiable as bots (which they would be if they were, for instance, marked as "bot"). These users don't want to use polibots to trick people about how many people support issue *x* and in *that* way convince them to support *x*. They want them to access the information tweeted and become influenced by considering that information seriously. Hence, the main intention and goal behind the act (of activating polibots) seem acceptable, even if effectiveness in reaching the goal (making people access the information and considering it seriously) might require using bots that aren't so easily identifiable as bots.

Of course, the opponent might respond that this nevertheless means that my use of polibots *relies* on tricking people into thinking the bots are real. My plan relies on fooling others even if my ultimate goal seems fine. Hence, it seems that I cannot be completely free of the charge that I am deceiving my peers. In response, however, I think we should draw a relevant distinction. This is the distinction between a plan's relying on *tricking* others into thinking that my bots are genuine humans, and a plan's relying on others not identifying my bots as bots. Take the person above who wants to convince others that *x* is correct by making others believe their pro-*x* bots are humans and in *that* way influence their opinions about x; and compare me, who want people to access the information shared by my bots only and in that way influence their beliefs. Both of our plans, in a sense, relies on people not identifying our bots as bots, but for very different reasons. My plan does so because, given the propensity of humans to be skeptical of bots and their content, my intended goal depends on them not identifying the bots. Call this sort of reliance on false beliefs for *negative* reliance. It is only a reliance on a false belief because the opposite of the false belief will thwart my aim. My plan only relies on anonymity of bots because of people's prejudice against bots. The other person, however, will *positively* rely on false beliefs in the sense that their entire plan is to deceive and in *that* way influence. For them, false beliefs are the means; for me, false beliefs are only contingent, unfortunate requirements. In a sense, the relevant distinction is between contingent and necessary reliance on false beliefs; and the anonymity of the bots serve very different functions in our respective plans—in one case as the means of getting others to support x, in the other case merely to avoid triggering scepticism and counter-productive effects. To me, a plan that only has negative reliance on false beliefs in this way doesn't seem so obviously deceptive.

Moreover, there is always also a separate question, namely about permissibility

*all things considered. Even* if *Personal Polibot* use would constitute some form of deception, we can ask how wrong, *all things considered*, that is. As will become clearer below, there might be countervailing reasons. For the main intention, as I said, is to get people to access information believed to be genuine and relevant to political matters. This is exactly the sort of information and method we *want* people in a democracy to base their opinions on. Hence, if polibots (often) can help boost this method, then the goodness (or democratic value) of this might outweigh the badness of any deceptive side-effects.

### 3.4. Objection 4: Polibots Impermissibly Influence Opinions.

A worry very similar to the one above is the following. An opponent might say that whether or not we call it deception, the problem with polibot use is this: 1) People are influenced by perceived popularity; 2) using polibots contributes to creating false impression of popularity around candidates and issues; 3) hence, polibots influence political opinions in a deceitful way. Woolley and Guilbeault (2017) have already discussed this sort of phenomena which they call "manufacturing consensus". One can make it seem as if many, many people agree with your beliefs, and in that way influence others' beliefs. Above, I tried to suggest that this won't be the primary intention of those using Personal Polibots. The worry here, however, is that even where such effects are unintended, it is nevertheless morally objectionable if polibots do have these effects. Before considering this objection, we should get clearer on how the problematic effect in question can be produced by polibots.

One direct way is if there are two political candidates, for instance, C and D where D is more popular. If I had control over many polibots and made these share pro-C sentiments, it will become more likely for another person online to encounter a C-supporter than a D-supporter. Over time, this could make others come to believe that C is *actually* more popular than D. In turn, this might influence the opinions that others have of C and potentially make them less confident in D or more confident in C. After all, they might think that all those C-supporters cannot be wrong. Admittedly, however, this effect is questionable. At least research done on the effect that *polls* have on political opinions seems to suggest that few people tend to *switch* sides based on perceived popularity.[8] Even so, since the number of actual supporters in many elections seem to be in the millions it seems unlikely that any single individ-

8.    See Snyder (2012).

ual will be able to have this effect on perceived popularity. It seems to be a worry that is more relevant when we consider the aggregate use *Personal Polibots* of many people.

However, it should firstly be noted that this sort of effect would not be particular to polibots at all. Indeed, news agencies might often unconsciously (or consciously, of course) achieve similar effects when they report more on certain candidates over others when this disproportionate attention to certain candidates have no basis in actual popularity. Similarly, if C's supporters are more active and effective in their canvassing efforts, they might get a similar effect on people even if D is actually more popular. Yet canvassing seems to be an activity we find permissible. Secondly, this feared effect is already built into social media and the internet itself because not all citizens are online in the first place. And there is no guarantee that the number of supporters for C and D online will be an accurate representation of the actual numbers either, for D might be more popular amongst more senior voters which tend to be less active online. Hence, basing one's beliefs about candidate popularity on social media is already risky, and the effect (false impression of popularity) is already likely to exist online whether or not polibots exist. Moreover, insofar as the opponent is criticizing the use of polibots for contributing to inaccurate representations of support online, he should also, by the same token, criticize those currently *not* online for doing the same. Yet, this seems like a path we do not want to go down. Third, as having an online presence online becomes more and more common for the entire population, and the polibot technology remains easily accessible, it might be the case that my 100 polibots in favour of C will be counterbalanced by 100+ polibots in favour of D; in which case the effect from polibots would disappear or decrease as the proportions of support would remain constant or at least similar. Lastly, there are also other, and more officially recognized, sources of candidate popularity, such as polls. Although not guaranteed to be accurate, these do make a serious effort to be as accurate as possible, whilst there isn't necessarily any such effort made by social media platforms. This should, hopefully, help balance false perceptions of popularity that might be created online. In sum, then, we can see that the problem of creating false impressions of popularity might not be such a persuasive issue for the bot use that we focus on, nor is it particular to polibots.

There is also, however, a less direct way of influencing opinions which Woolley and Guilbeault (2017) call "agenda setting". One can influence political opinions by controlling what topics and issues are discussed and perceived as important. This also seeps out into the real world, as news stations and others pick up on "hot topics"

online and report on that. Suddenly, it can seem to everyone that some issue is very important, even though it actually isn't. Further, if this issue puts C in a good light in some way, it can make people more inclined to support C as well. In sum, through affecting what is talked about and viewed as important, polibots can change opinions. Although there are admittedly frightening examples of this effect[9], the worry as it relates to *Personal Polibots* seem less serious. Indeed, it seems that many of the responses given above hold with equal force here. To reiterate, as well, it seems unlikely that a single individual (within the confines of Personal Polibots) will be capable of wielding such power. Hence, the agenda setting effected by polibots will largely emerge out of many distinct individuals using their own *Personal Polibots*.[10] It can be hoped that, as more and more would use the technology, one potential effect of this is that the sorts of issues that get set highest on the agenda will mirror the issues that people, in actuality, do put most weight on as well. It is, of course, always a problem that the issues and topics that people will boost with their bot uses will turn out to be fake or deceptive itself—perhaps due to the large amounts of fake news circulating. The problem, here, however is first and foremost a problem about fake news itself, and not necessarily about the polibots. Furthermore, it is also a possibility that the polibots would share and boost genuine news and in that way help countervail the influence of fake news online. Hence, it is very unclear what the effects will be. And it seems to me that, morally, the goodness or badness of this sort of influence on opinions will depend a lot on the source of the information shared and boosted as well.

Even if all the issues boosted were genuine and morally important, the critic might insist that there is a possibility that some will be influenced in the wrong way, i.e. *purely* by the perceived popularity of the case, not the substance itself. It bears mentioning however, that any sharing of information runs the risk of influencing opinions in suboptimal ways. This possibility seems almost impossible to get rid of, with or without, polibots. Moreover, this seems more of a problem with human psychology itself, and not directly the polibot use in question. Even doing our best to reduce this sort of influence, it seems inevitable that someone will be influenced by mere popularity.

Lastly, it is also worth mentioning some potentially positive uses of agenda

---

9.    Woolley and Guilbeault mention "Pizzagate", the conspiracy linking Hillary Clinton to human trafficking and child abuse, which seems to have been boosted and made into a much talked about issue online, partially, by the use of polibots.

10.    I am focusing here on aggregates of distinct individual uses.  When we start thinking about collectively acting groups and coordinated polibot campaigns, many distinct worries arise.

setting with polibots. Currently, some information about political candidates and certain issues often get less attention in the news, etc. for no "morally" good reason. A damaging report about D's conduct towards women might get little attention because, for instance, D is backed by powerful people who try to keep it out of the discourse, or D is very popular and his supporters do their best to keep the stories out of the online discussions by not engaging with the stories at all. Similarly, a morally and politically important issue might get little attention simply because it is a minority issue. None of these are morally good reasons for their unpopularity. It can be hoped, then, that individuals might be able to use polibot technology to bring such issues and information more to the forefront of discussions, and force people to engage with them more seriously. This seems to me a virtuous use of the technology. Moreover, at least for many realistic such scenarios, it seems as if the desired effect isn't merely to sway political opinions by sheer perceived popularity even though that can be an effective means for, say, an underrepresented group. More centrally, I think, many in such a position would not simply aim to change policies by making politicians believe there is massive support for it, but more directly through getting these politicians (and others) to have a change of heart themselves. In other words, getting people to *realize* that the issue is real and important, and in *that* way get other genuine humans to be morally motivated to support a change. They would seek a genuine recognition of the problem; and that requires changing hearts, not merely pushing through unmotivated change.

In sum, then, I believe the objection is overestimated. I've tried to argue that, insofar as the problem is potentially real, it will not necessarily be widespread on the individual level. Nor is the problem unique to polibots either. Lastly, there also seems to be potentially good uses of the effect in question.

*3.5. Objection 5: Collective Polibot Use Creates Chaos and Distrust.*[11]

Much of the discussion so far have focus on the individual use of *Personal Polibots*. Let us move instead to more large-scale effects that would potentially emerge if many, or indeed most, started using *Personal Polibots*. There might be a worry that chaos and distrust will increase. Suppose the two candidates C and D have 1 million and

11.　This objection is inspired by some brief remarks made by Julian Savulescu during the presentation of a shorter version of this paper in the final round of the 2017 Oxford Uehiro Prize in Practical Ethics competition.

5 million actual supporters online respectively. Suppose that each supporter would use 100 polibots each. Suddenly we go from a pool of 6 million users to, effectively, a pool of 300 million users. That is an extreme increase. It is easy to imagine that chaos would ensue, with the social media platform being dominated by mere polibots talking past each other. As this increases, people will find it harder to find other genuine humans. They will start becoming much more distrustful of users they meet online as they know it is most likely a fake one. As this increases, people will also likely start to look for other less bot-infested platforms and move there, which again will become infested, and people will migrate elsewhere again, and so on. Indeed, polibots might undermine the essential function of social networks: being a virtual space for people to discuss and engage with each other.

Despite the dystopian feel to this scenario, I believe it would most likely be averted in a natural way. For instance, just like websites limit adverts and email services limit spam emails, it could be imagined that companies like Twitter and others would regulate the total number of polibots permitted on their sites at any one time, or would attempt to regulate in some non-biased way the number of polibots that are *shown* in people's feeds. It is also likely that user themselves might act so as to decrease the "chaos". We are all already familiar with "blocking" technology, in relation to adverts and spam emails. It seems to me likely that "bot-blocking" technology that decrease the number of bots visible in one's feed would emerge as the use of Personal Polibots increase as well. Moreover, it might be that such a natural response is to be preferred over a desire for a complete eradication of *Personal Polibots*.

The reason is this. Most blocking technologies are seldom 100% effective, as people find different ways to trick them. It is likely therefore that some polibots will get past a person's blockers in any case. This is important as most people online do find themselves in self-imposed echo chambers and informational or filter bubbles. [12] These embed us in social media feeds containing information the algorithms believe we already approve of. This is a democratically unfortunate effect as people are less likely to be confronted with opposing information, arguments and views. If the polibots outside these bubbles are many enough, however, it is more likely that some will seep through the both the filter bubble and the blockers, and thereby help confront people inside them with relevant information that wouldn't otherwise easily come across. After all, even inside a blocking-enforced bubble it is hard to hide from very popular and widely spread issues. And the more users appear to discuss a certain

12.   See Hossain (2016)

issue, the more likely it is that one of these users—whether genuine or a bot—will slide into one's feed, past the bubble and the blockers.

In sum, then, I believe there might be responses that emerge naturally to combat any feared chaos and distrust online. These natural responses might moreover be favourable to a complete eradication of the bots themselves. The more people are talking about a certain issue and using polibots to boost that discussion, the more likely it is that others will be confronted with those issues (without encountering the chaos), even hiding behind bubbles and blockers. Indeed, the result might be that people are more exposed to the information that matters the most.

*3.6. Objection 6: Permissible Use of Polibots is Practically Impossible.*

As we've seen throughout the paper, I have only defended a very limited type of use of polibots. Plausibly, we might require that bots do not share hate-speech, do not spread deceitful information, isn't used merely to suppress others, and so on. Potentially, we might also need some form of regulation to avoid that some people get disproportionality many polibots. However, it seems very hard to ensure that these requirements will be obeyed. To ensure this, we need technology that is capable of identifying both bots (and who controls them) and their content. Yet, this will be extremely difficult to achieve. Hence, fearing that "bad bots" will thrive unless we have this technology, it is overall more advisable simply to discourage the use of polibots, and indeed work to shut all of them down. This objection, then, accepts the claims of this paper but argues that there is no way of practically allowing these sorts of bots without having an increase in bad forms of bot use as well.

The problem with this argument is that it is overly pessimistic about technology. For instance, the fact that researches are able to analyse the numbers and activities of polibots, and indeed their origin, shows that we already do have some methods available to detect bots. Moreover, we already know that companies like Google and Facebook are already developing automated technologies that can detect fake news, hateful comments, and so on.[13] It doesn't seem to me, therefore, that technology capable of ensuring compliance with the requirements on polibots like the ones mentioned above is too hard to engineer.

13.    See, e.g., Feldman (2017), Leong (2017), and Frier (2017).

## 4. CONCLUSION

This paper has attempted to highlight and discuss some initially intuitive worries regarding one potential use of polibots, namely the personal use of polibots to enhance one's political voice. Although this discussion has helped clarify some ethical issues, all of the putative objections suffer from weaknesses. Many of them are very general and would equally apply to technologies we do find permissible; some of them clearly fail; and some of them are very likely overestimated in their force. Moreover, I've also tried to highlight some potentially positive and valuable effects of polibot use.

In sum, I've tried to outline a plausibly permissible use of polibot technology which I think escapes much of the intuitive worries we might have about them. At the end of the day, I fear that most of the scepticism towards this use might arise because of the word "bot" itself. As we look deeper down at the potential problem areas, however, we see more clearly to what extent this use of the technology is just a new way of speaking up in today's world—the kind of technology we've sought since the beginning of democracies, and even before that.

However, we should keep in mind that although I have played the devil's advocate on behalf of *Personal Polibot* use as a mere extension of one's political voice and agency, there are many other areas concerning different polibot uses (such as those by governments and official campaigns), related activities (such as the commercial selling of polibots), future polibot technology (such as more autonomous ones) that deserve their own ethical examinations.

## REFERENCES

Alfonso, F. (2015). *Twitter Bots Silence Critics of Mexico's Leading Presidential Candidate*, from The Daily Dot: https://www.dailydot.com/news/pena-nieto-twitter-bots-mexico-election/ [Accessed 25 July 2017].

Andrews, C. (2017). *Are Political Social Media Campaigns a Threat to Democratic Elections?*, from E&T: https://eandt.theiet.org/content/articles/2017/05/are-political-social-media-campaigns-a-threat-to-democratic-elections/ [Accessed 17 May 2017].

Bessi, A., & Ferrara, E. (2016). Social Bots Distort the 2016 U.S. Presidential Election Online Discussion. *First Monday*, 11(21).

Cook, J. (2011). *Update: Only 92% of Newt Gingrich's Twitter Followers are Fake*, from Gawker: http://gawker.com/5826960/update-only-92-of-newt-gingrichs-twitter-followers-are-fake   [Accessed 25 July 2017].

Feldman, B. (2017). *Can Google Use AI to Fix the Comment Sections?*, from NY Mag: http://nymag.com/selectall/2017/02/google-introduces-perspective-a-tool-for-toxic-comments.html  [Accessed 13 August 2017].

Finley, K. (2015). *Pro-Government Twitter Bots Try to Hush Mexican Activists*, from Wired: https://www.wired.com/2015/08/pro-government-twitter-bots-try-hush-mexican-activists/  [Accessed 25 July 2017]

Frier, S. (2017). *Facebook Automates Effort to Flag 'Fake News' for Fact Checking,* from Bloomberg: https://www.bloomberg.com/news/articles/2017-08-03/facebook-automates-effort-to-flag-fake-news-for-fact-checking [Accessed 13 August 2017]

Guilbeault, D., & Woolley, S. (2016). *How Twitter Bots are Shaping the Election*, from The Atlantic: https://www.theatlantic.com/technology/archive/2016/11/election-bots/506072/  [Accessed 25 July 2017].

Hallemann, C. (2017). *Here's What You Need to Know About Donald Trump's Lawsuits*, from Town and Country Magazine: http://www.townandcountrymag.com/society/politics/a9962852/lawsuits-against-donald-trump/ [Accessed 13 August 2017]

Harn, A. (2017). *Facebook and Twitter are Being Used to Manipulate Public Opinion*, from The Guardian: https://www.theguardian.com/technology/2017/jun/19/social-media-proganda-manipulating-public-opinion-bots-accounts-facebook-twitter [Accessed 25 July 2017].

Hossain, I. (2016) *Filter Bubbles are Shrinking Our Minds*, from Huffington Post: http://www.huffingtonpost.in/ishtiaque-hossain/filter-bubbles-are-shrinking-our-minds_a_21469747/  [Accessed 27 August 2017].

Lee, T. Y. (2017). *Bots Used to Bias Online Political Chats*, from BBC News: http://www.bbc.com/news/technology-40344208 [Accessed 25 July 2017].

Leong, L. (2017). *Fighting Fake News: How Google, Facebook and Others are Trying to Stop It*, from Tech Radar: http://www.techradar.com/news/fighting-fake-news-how-google-facebook-and-more-are-working-to-stop-it [Accessed 13 August 2017].

Lynch, M. (2016). The Internet of Us: Knowing More and Understanding Less in the Age of Big Data. New York: Liveright.

Mahon, J. (2008). *The Definition of Lying and Deception*, from Stanford Encyclopedia of Philosophy: https://plato.stanford.edu/entries/lying-definition/ [Accessed 25 July 2017].

McKelvey, F., & Dubois, E. (2017). Computantional Propaganda in Canada: The Use of Political Bots. Working Paper 6, Oxford University, Project on Computantional Propaganda.

Miller, C. (2017). *Governments Don't Set the Political Agenda Anymore, Bots Do*, from Wired: http://www.wired.co.uk/article/politics-governments-bots-twitter [Accessed 25 July 2017].

Newman, H., & O'Gorman, K. (2017). *Political Bots are Poisoning Democracy - So, Off With Their Heads*, from The Conversation: https://theconversation.com/political-bots-are-poisoning-democracy-so-off-with-their-heads-79779 [Accessed 25 July 2017].

Snyder, B. (2012) *How polls influence behavior*, from Stanford Graduate School of Business: https://www.gsb.stanford.edu/insights/how-polls-influence-behavior [Accessed 24 january 2017].

Woolley, S. (2016). Automating Power: Social Bot Interference in Global Politics. *First Monday*, 4(21).

Woolley, S., & Guilbeault, D. (2017). *Computantional Propaganda in the United States of America: Manufacturing Consensus Online*. Working Paper 5, Oxford University, Project on Computational Propaganda.

# What Makes Discrimination Wrong?

PAUL DE FONT-REAULX

*University of Oxford*

2017 OXFORD UEHIRO PRIZE IN PRACTICAL ETHICS

WINNER: UNDERGRADUATE CATEGORY

## ABSTRACT

Most of us intuitively take discrimination based on gender or ethnicity to be impermissible because we have a right to be treated on the basis of merit and capacity rather than e.g. ethnicity or gender. I call this suggestion the Impermissibility Account. I argue that, despite how the Impermissibility Account seems intuitive to most of us with a humanist outlook, it is indefensible. I show that well-informed discrimination can sometimes be permissible, and even morally required, meaning we cannot have a strict right not to be discriminated against. I then propose an alternative and more plausible account which I call the Fairness and Externalities Account, arguing that acts of discrimination are wrong partly because they are unfair and partly because they create harmful externalities which—analogously to pollution—there is a collective responsibility to minimize. Both of these factors are however defeasible, meaning that if the Fairness and Externalities Account is correct, then discrimination is sometimes permissible. These results are counterintuitive, and suggest that the ethics of discrimination requires further attention.

## 1. WHAT IS DISCRIMINATION?

I take discrimination to be to treat someone very differently in an unfavourable way based on an irrelevant trait. A trait is relevant if and only if the possession of it by itself provides reasons for different treatment in some instance, such as constituting a difference in merit or capacity. Otherwise it is irrelevant. For example, choosing members of a sports team based on athletic ability is not an instance of discrimination, because athletic ability is a relevant trait for being a member of a sports team. Doing the same on the basis of ethnicity is however, because ethnicity is not a relevant trait[1].

Discrimination caused by bigotry such as racism is often indefensible simply because it rests on ungrounded beliefs about the relevance of traits such as ethnicity, such as the belief that a given ethnicity is relevantly superior in some normal situation. Discrimination can be wrong even if it does not suffer from epistemic problems however. This is when different treatment is based on an irrelevant trait, but there are good epistemic reasons—such as statistical evidence—to believe that holding this trait makes it more likely that the same person holds another relevant trait. For example, ethnicity is not a relevant trait for performing a normal job. Criminality is however, and for various reasons (e.g. social segregation along ethnic lines) ethnicity can statistically correlate with criminality. In such cases, discriminating on the basis of ethnicity constitutes an instance of what I will call 'epistemically grounded discrimination', which is what I will focus on in this text.

## 2. IS DISCRIMINATION IMPERMISSIBLE?

Many of us believe that even epistemically grounded discrimination is sometimes wrong, such as in some cases of ethnic targeting by police. I will now progress to consider an account of under what conditions an instance of epistemically grounded discrimination is wrong, and what it is that makes it so.

---

1.    We might use the word 'discrimination' as simply referring to different treatment based on some trait. I will reserve the term for this 'problematic' sense, requiring selection based on an irrelevant trait. I also believe this definition corresponds closely to how most people use the word.

Those of us with a humanist outlook often have a *prima facie* intuition that discrimination is always wrong in any realistic scenario. Here is an account of the wrongness of discrimination which corresponds to this intuition: We have the *right* to be judged based on individual merit and capacity rather than generalizations over traits for which we are not responsible. Specifically, all individuals have a right not to be judged or treated *differently based on traits which do not constitute a relevant difference in merit* (e.g. having committed a crime) *or capacity* (e.g. athletic ability). Furthermore, it is impermissible to violate someone's rights. Discriminating based on e.g. ethnicity or gender is to violate someone's rights in this way. For this reason it is impermissible. Let us call this the *Impermissibility Account*.

As mentioned, I believe the Impermissibility Account is at least *prima facie* intuitive to many of us. This makes it worth studying in more detail. Let us test the suggestion by considering the following two examples:

Hiring: Harold is considering applicants for a position at his company. He knows that being a member of an ethnic minority strongly correlates with frequency of crime. Because of this he chooses not to consider applicants belonging to an ethnic minority.

Au Pair: Cassandra is considering hiring an au pair to take care of her children. She knows that being white and male strongly correlates with being inept at taking care of children. Because of this she chooses not to consider white male candidates.

These are both instances of epistemically grounded discrimination, and impermissible according to the Impermissibility Account. In their respective examples Harold and Cassandra both treat some group of individuals very differently based on an irrelevant trait (being of an ethnic minority and being a white male respectively). They do however have good reason to believe that these traits correlate with a relevant trait (criminality and child-caring respectively), and that therefore—ceteris paribus—someone with the irrelevant trait (e.g. being a white male) is more likely to possess the relevant trait (e.g. being inept with children) than someone lacking the irrelevant trait.

Most of us believe that Harold is acting impermissibly, and plausibly for the reasons above. He is treating individuals in a way which they have a right not to

be treated by discriminating based on ethnicity. Many of us do not however find Cassandra's action impermissible, and feel inclined to be more lenient in her case. In fact many ads for au pairs specify that they only consider females, and the owners of the ad are rarely considered to be doing something impermissible. This would suggest that our suggestion is too inclusive, making too many acts of discrimination impermissible. In order to still be able to consider Harold's actions impermissible however, we need to find some further necessary condition which only rules out Hiring.

We might suggest that the relevant difference between Hiring and Au Pair are the objects of discrimination. Here is a further condition attempting to accommodate this suggestion: Discrimination is only impermissible when the irrelevant trait (e.g. being a white male) constitutes the membership of an *exposed group*. By 'exposed' I mean a group which is generally considered to be at the receiving end of structural oppression, such as ethnic minorities or women[2]. Cassandra's discrimination is directed at white males, which are not an exposed group, while as Harold's actions are directed at an ethnic minority, which is an exposed group. Therefore Harold's discrimination is impermissible, but not Cassandra's.

Adding this condition to the Impermissibility Account captures a further intuition, but the condition is both problematically vague and insufficient. It is problematically vague because there are cases in which it is not clear whether a trait should warrant membership in an exposed group or not (e.g. having one grandparent of an ethnic minority). I will not elaborate on this issue. More relevantly it is insufficient because sometimes it seems permissible, and even morally required, to discriminate even when the object of discrimination is a member of an exposed group. Consider the following example:

> Alley: In order to get to the other side of town one needs to pass through either alley A or alley B. A tourist, who is internationally famous for spreading money around everywhere he goes, needs to get to the other side of town, and asks Pasha how to get there. Pasha knows only the following: (a) Passing through A is slightly quicker than B, (b) A is mainly populated with members of an ethnic minority and B by non-minorities and (c) being a member of the ethnic minority statistically correlates strongly with high

---

2.    What it means to be at the receiving end of structural oppression is a complicated question. Since this is not an account I intend to defend here I will simply assume whatever interpretation of it that is the most charitable.

frequency of violent crime, while there is no such correlation for non-minorities. The tourist will only understand a simple direction to A or B, and will then trust Pasha's recommendation blindly.

In this case it seems to me that Pasha is not only permitted to direct the tourist to B despite A being the quicker path, but is also morally required to do so. The tourist is left at Pasha's mercy, and directing him to A would likely put him at great risk. Doing so would however imply discriminating against the individuals in A which are members of an exposed group, because Pasha would treat them very differently by recommending the tourist to avoid them only on the basis of their ethnicity, thereby robbing them of the chance to receive some of his significant spending[3]. If this is correct, it cannot be impermissible or even always morally wrong to discriminate on the basis of irrelevant traits which constitute membership of an exposed group. This implies that the Impermissibility Account cannot be correct, and that we should reject it as an account of the conditions under which epistemically grounded discrimination is wrong. It follows that discrimination based on e.g. ethnicity is sometimes permissible.

## 3. FAIRNESS AND EXTERNALITIES

We might find it surprising that we cannot defend the Impermissibility Account. This does not mean however that epistemically grounded discrimination is always, or even often, permissible; most often it is not. To determine when this is the case we need an alternative—*defensible*—account of under what conditions an instance of epistemically grounded discrimination is wrong, and preferably of what it is that makes it so. Furthermore we would like such an account to capture variations in the wrongness of discrimination depending on the objects of discrimination.

I suggest that the best account of the wrongness of discrimination, and the one that we should prefer over the Impermissibility Account, is what I will call the *Fairness and Externalities Account*. This account states that the wrongness of discrimination is constituted by two independent and defeasible factors:

---

3. Notice that this example is structurally similar to Hiring, in that both Harold and Pasha *avoid* treating some individuals based on an irrelevant trait. For this reason we should not object only to Alley as an example on the basis that Pasha does not 'treat' the inhabitants of alley A, because on that notion Harold does not 'treat' his applicants either.

*Fairness*: Discriminating against an individual is *unfair* to that individual, and all else equal it is wrong to treat an individual unfairly.

*Harmful Externalities*: Discriminating against an individual produces *harmful externalities* for individuals sharing the trait which has been the grounds for discrimination.

Fairness is something like being treated in proportion to one's desert, need and capacity[4]. When someone is not selected for a job because of their ethnicity, this is unfair to them, because they have *ceteris paribus* equal desert, need and capacity to perform the job to other applicants. For this reason both Harold and Cassandra are treating their respective applicants unfairly. To be treated fairly is however not an unalienable right, and sometimes we are permitted to treat people unfairly when it is required for overriding reasons. An example of this is might be collective punishment of a platoon for the mistake of an individual, when it is required for discipline. When there are no such overriding considerations it is wrong to treat individuals unfairly.

I believe the fairness condition captures something intuitive about the wrongness of discrimination, in that we all feel—no matter our background—like we are treated unfairly when someone behaves differently to us in an unfavourable way due to some irrelevant trait of ours. I also believe that it is plausible to say that it is *ceteris paribus* wrong—when nothing sufficiently important is at stake—to treat someone unfairly in such a way. This condition seems sufficient to explain why all instances of discrimination seem at least somewhat wrong, including e.g. Cassandra's. It does not however explain why we intuitively believe that it is much worse to discriminate against some individuals and groups than others, e.g. why Harold's act of discrimination seems much worse than Cassandra's. To do this we need to consider the second condition.

Acts of discrimination can be harmful, e.g. if one of Harold's applicants is very poor and in need of a job. One might suggest that the harm caused to the objects of discrimination is larger in Hiring than in Au Pair, and that this is what makes Harold's actions worse than Cassandra's. However, variation in the harm caused to the individual being discriminated against is insufficient to account for the variation

---

4.   Because fairness is a complicated and contested concept I believe it is wiser not to attempt to give a precise definition in this essay, at the risk of diverging from the main discussion. Instead I rely on the assumption that the reader sufficiently shares my intuitive conception of fairness.

of wrongness that we find depending on the object of discrimination. For example, it might be that some of Harold's applicants are very well-off and not significantly harmed by being discriminated against. Rather there seems to be something relevantly wrong about the discrimination being based on ethnicity, independently of the harm caused to the discriminated individual.

I argue that this can be explained in terms of *harmful externalities*. An externality is '*a cost or benefit that affects a party who did not choose to incur that cost or benefit*'[5], which includes but is not limited to the harm caused to the object of discrimination. When someone is discriminated against on the basis of a specific trait, this can produce harmful externalities not only for them but also for others sharing that trait, e.g. by reducing opportunities, perpetuating stereotypes or solidifying segregation along ethnic lines. For example, when Harold systematically disregards an exposed group he is signalling to other employers that this group is not to be trusted, making them less likely to hire members of the ethnic minority. This causes a higher level of unemployment among the ethnic minority and perpetuates alienation, which constitutes a harm to members of that group. In other words, the harm act of discrimination can extend beyond the target of discrimination to others who share her traits.

I believe this can account for cases like Hiring and Au Pair. For example, I would argue that the harm caused to white males by Cassandra by e.g. limiting the group's chances to improve its child-caring skills is non-negligible, but comparatively limited. Harold on the other hand might more plausibly be contributing to a deep societal problem of segregation along ethnic lines, which is much more problematic. This condition can account for why Harold's actions are wrong while Cassandra's are plausibly permissible, because discriminating against an ethnic minority produces more harmful externalities than discriminating against white males.

I believe the Fairness and Externalities Account captures the wrongness of discrimination in an intuitive way by both capturing why we think that there is always something wrong about discrimination, but also why the wrongness can vary with the object of discrimination. Accepting the account would however have some counterintuitive consequences:

Firstly, the nature of externalities is often such that they affect large groups and are produced collectively by many individuals, as is the case when it comes to pollution. The responsibility not to produce such externalities is collective, and any individual responsibility is a part of that collective responsibility. This means that ac-

---

5.   Buchanan & Stubblebine (1962)

cording to the Fairness and Externalities Account the wrongness of discrimination is largely made up of a failure to live up to one's part in the collective responsibility not to e.g. perpetuate alienating stereotypes. Specifically, any further wrongness to discriminating against an exposed group relative to e.g. white males is fully constituted by one's marginal contribution to negative externalities for that group. This is counterintuitive, because discrimination is intuitively thought of wronging someone, not as—like driving a car with unnecessarily high emissions—contributing to the production of negative effects for a group.

Secondly, like with pollution there can sometimes be overriding reasons to disregard this collective responsibility when the marginal harm produced is significantly outweighed by the harm of alternative actions. This is what happens in Alley, where the risk to the tourist outweighs the unfairness and harm caused to the individuals in alley A. Because both Fairness and Harmful Externalities are defeasible, discrimination is permissible when the unfairness and negative externalities are counterweighed by independent considerations. In Cassandra's case this seems quite intuitive a well: if the probability of a white male applicant being good with children is significantly lower than other candidates, this might plausibly outweigh the unfairness and harm caused to the white male applicants. On the other hand however, it seems then that Harold might also be allowed to discriminate based on ethnicity if, for example, his business was extremely sensitive to criminality and the correlation between ethnicity and criminality sufficiently strong, which appears counterintuitive.

## 4. CONCLUSION

I have argued that unalienable rights are not sufficient to explain the wrongness of discrimination by showing why the Impermissibility Account is implausible, and that it should instead be explained by unfairness and negative externalities in accordance with the Fairness and Externalities Account. If this is correct then this means that we should shift our view of the wrongness of discrimination to it as largely being a failure of a collective responsibility, and that we have to consider for each case whether there are sufficient independent considerations to make discrimination permissible. For example, how sensitive does Harold's business need to be to allow him to discriminate? I believe this is counterintuitive, and indicates that the ethics

of discrimination is more complicated than it might first appear and deserves further attention[6].

*Acknowledgements*

*Several people have given very helpful comments on the content of this text. For this I'd like to thank Alex Brian, Thomas Douglas, David Edmonds, Jeppe Egsgaard, Jacob Lagerros and Viking Waldén. Furthermore I'd like to thank Jeff McMahan, without whom this text would not have been written in the first place.*

REFERENCES

Alexander, Larry (1992) *What Makes Wrongful Discrimination Wrong? Biases, Preferences, Stereotypes, and Proxies*. University of Pennsylvania Law Review 141:1, pp. 149-219.

Buchanan, James & Stubblebine, Craig (1962) Externality. *Economica* 29:116, pp. 371-381.

Edmonds, David (2006) *Caste Wars: A Philosophy of Discrimination*. New York: Routledge.

Fricker, Miranda (2007) *Epistemic Injustice: Power and the Ethics of Knowing*. New York: Oxford University Press.

Hellman, Deborah (2008) *When is Discrimination Wrong?*. Cambridge, MA: Harvard University Press.

Lippert-Rasmussen, Kasper (2014) *Born Free and Equal? A Philosophical Inquiry Into the Nature of Discrimination*. New York: Oxford University Press.

---

6.    Due to the format of this essay it was not written with the intention to respond to previous literature on the topic. My arguments, I have later discovered, parallel some by Lippert—Rasmussen (2014) who also focuses on the harm of discrimination. Other important entries in the ethics of discrimination include Alexander (1992), Edmonds (2006), Fricker (2007) and Hellman (2008).

# The Ethical Dilemma of Youth Politics

ANDREAS MASVIE

*University of Oxford*

2017 OXFORD UEHIRO PRIZE IN PRACTICAL ETHICS

RUNNER-UP FINALIST: UNDERGRADUATE CATEGORY

ABSTRACT

Youth politics is celebrated locally, nationally and internationally. In this paper, I argue that such celebration must be deemed unethical. First, I assert that the vitality and sustainability of any democracy is contingent upon the people's ability to reason and think critically. Second, I make the case that youth politics exerts negative effects on the members' ability to reason and think critically. Then I approach Norway as a case study, exemplifying how these negative effects may play out. Thus one is faced with the following dilemma: It is regarded ethically imperative to promote political engagement amongst youth, but it is unethical to promote youth politics.

The West in general, and perhaps Europe in particular, tends to celebrate youth politics as a vital force of democracy. This is reflected in the current literature on youth politics, which appears to be almost exclusively descriptive (e.g. 'What is the level of youth politics in country x?') or positively normative (e.g. 'How can country x heighten engagement in youth politics?'). Various youth councils and parliaments are encouraged and empowered by government as well as civil society, both at local and national level. This is also the case internationally. The UN, for instance, demands that youth politics be stimulated: "[Such] engagement and participation is central to

achieving sustainable human development.["] I will approach the rationale of this collective celebration as a syllogism, defining 'youth politics' as organized political engagement of people aged 13—25:

> P1   Youth politics increases the level of political engagement;
> P2   Political engagement promotes democratic vitality and sustainability; *thus*
> C1   Youth politics promotes democratic vitality and sustainability.

In this paper I am interested in challenging P2. Does the increased political engagement due to youth politics promote democratic vitality and sustainability? For the sake of argument, I will posit the trueness of P1. When it comes to P2: It would be difficult to argue that all forms of political engagement promote democratic vitality and sustainability (e.g. authoritarian neo-Nazism or revolutionary Communism). Hence, I shall take it for granted that P2 is constrained to activities and policies compatible with democracy.

I intend to demonstrate that youth politics exerts a negative influence on democratic vitality and sustainability, and thus, that it is unethical to continue the current celebration of youth politics.  This requires doing four things: In section one, I will develop a methodology for evaluating whether something promotes democratic vitality and sustainability or not. In section two, I will apply this methodology to youth politics. In section three, I will present empirical indicators from Norway, highlighting some evident problems of youth politics. Finally, in section four, I will sketch solutions to these problems.

## I. METHODOLOGY FOR EVALUATING EFFECTS ON DEMOCRATIC VITALITY AND SUSTAINABILITY.

How does one evaluate effects on democratic vitality and sustainability (henceforth simply 'DVS')? First, I will assume that there exists some objective theory of DVS, though I will not attempt to detail any such theory here. Second, I will assume that any normally functioning person is able to rank a selection of democracies according to their DVS, given that he or she is provided with sufficient data (e.g. crime and corruption rates): even though one might not agree on the exact details of the

---

1.   One of UNDP's "strategic entry points to enhance youth engagement and participation" is to strengthen "youth political participation at all levels." Bureau for Development Policy. (2014).

objective theory of DVS, shared intuitions enable different persons to identify vital and sustainable democracies. Third, I will assume that any given democracy seeks to maximize its potential DVS.
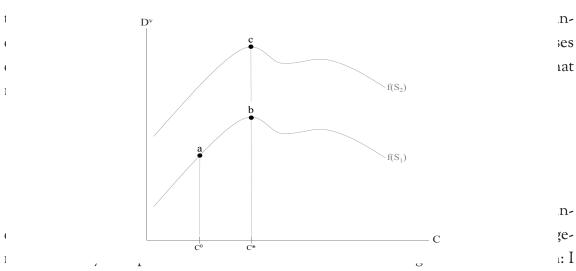
My method is simply to focus on the moment of election, these critical junctures of democracies. To a significant extent, the DVS is determined at such junctures. And here, at the junctures, the democracy faces the problem of maximizing DVS. I assert that maximizing DVS is contingent on rationality *qua* reason and critical thinking. Thus, in order to evaluate the effects on DVS, I shall isolate the mechanisms of rationality. In any given election, there are two such mechanisms at work. First, the more reasonable and critically thinking the *voters* are, the more likely they are to choose the candidate who produces maximum DVS, given the set of electable candidates. In short, reasonable and critically thinking voters will be better equipped to identify the most suitable of the electable candidates. Second, the more reasonable and critically thinking the *candidates* are, the more DVS they will produce. In short, the candidates will be better equipped to govern in a way that promotes democratic vitality and sustainability.

Any method must simplify reality in order to isolate significant factors, but why isolate reason and critical thinking? Reason enables the voters and candidates to move consistently from a set of premises to their conclusions. Reason also captures general comprehension as well as the aptitude for orderly thinking. Some of the premises that reason builds upon will be grounded in ideological axioms. Critical thinking enables voters and candidates to assess such ideological axioms and their application in the current context. A critically thinking person must be open to adjust axioms according to new knowledge. It does not imply, however, that his or her convictions move towards some fictitious notion of the 'perfect set axioms.'

Perhaps the emphasis on reason and critical thinking is justifiable for the voter, whose problem is to identify the most suitable candidate. But surely, some might infer, political virtues like compassion or prudence must, at least to *some* degree, affect the DVS that the candidate is likely to achieve? Most definitely—but the efficacy of all such virtues is contingent on reason and critical thinking. For instance, for a political candidate, it is never enough simply to exhibit compassionate intentions. To be of political value, compassionate intentions must materialize in the form of actions and policies that exert real impact on the relevant people. And the movement from 'compassionate intentions' to 'real impact' is to a large extent determined by reason

and critical thinking. The candidate must ask and answer questions such as: What effects ought to be realized? What measures do actually realize these effects?

For $P_2$ to be true, then, one of two conditions must befall. Either youth poli-

$D^v$

$c$

$f(S_2)$

$b$

$f(S_1)$

$a$

$C$

$c^0$ $c^*$

will argue that youth politics exerts two negative interrelated effects on DVS, those of *camaraderie* and *conformity*.

The first negative effect is related to camaraderie. When youths are introduced to politics, this usually happens through the youth wing of a political party. Which party one gets involved with is determined by the preferences held at that particular age, naturally heavily influenced by parents and friends.[2] Thus, it is arbitrary what political community the youth gets engaged with. I use 'arbitrary' in the sense that it is not a choice based on reason and critical thinking. It obtains, however, that reason and critical thinking cannot be achieved in a vacuum. Therefore, if the person does not engage in youth politics, he or she will certainly be influenced by other normative forces. Why should this be more desirable? The profound problem with youth politics is that the members develop relationships which inevitably depend on loyalty to a particular ideology and set of policies—the very fabric of the political community is political loyalty. In such a community, it seems challenging to pursue reason and critical thinking. This can be illustrated by the way in which parties tend to equate certain policies with the political ideology itself. I believe, for instance, this explains why some parties are so slow to re-evaluate their policies on drug criminalization. A simplified account of their rationale might be: people suffer from drug abuse; there is less drug abuse if drugs are criminalized; therefore drugs should stay criminal-

2. A Norwegian report suggests that the influence of parents might be particularly strong: "Det vi vet er at den politiske sosialisering i hjemmet har stor betydning for de unges deltakelsesmønster." Ødegård & Bergh (2011).

ized. But an accumulating body of empirical evidence demonstrates that the second premise is unjustified. That is, it is not necessarily the case that there will be less drug abuse if drugs are criminalized.[3] Why do not the parties adjust their conclusion? Perhaps they simply do not believe that the empirical evidence refutes the second premise. However, the lack of willingness to properly re-evaluate their conclusion suggest that the parties do not seriously take into account the new evidence. Why? One reason might be that the parties have made polices on drug criminalization an integrated aspect of their ideological core. This would imply that the members of the party do not only have to accept such irrationality, but must also internalize it: members might argue and disagree on the party's peripheral policies, but it is difficult to challenge the very core of the party's ideology if one is to remain in good standing within the party. Challenges against the core of the party's ideology inevitably pose an existential threat to the community. Furthermore, because of camaraderie, there are also barriers to leaving the community, as it entails leaving a social community that partakes in identity formation.

The second negative effect is related to conformity, and comes as a consequence of camaraderie. The fabric of the political community is, as already mentioned, political loyalty. Hence, a political community naturally develops mechanisms for defending its ideology against challenging arguments, rather than evaluating the counter arguments on their own merit. This is a problem related to all sorts of party politics. Some might even argue that youth wings are able to initiate changes in the parties (e.g. on drug policies), and thus acts against conformity. And this might indeed be true, to some very limited extent. But this misses the most important point: in order to progress within the party, one has to conform to the party line. That is, members desiring political careers within the party must conform to certain views and adopt them as their own, which is a process that is initiated in the youth wing. After all, if the person does not conform, why should he or she progress to represent the party at large? Those who hold and favor views in line with the community as a whole rise through the ranks, gain more influence over the community, and thus strengthen the effect of conformity. The negative effects on reason and critical thinking are obvious.

I cannot think of any massive gain in political virtues that might justify the negative effects of camaraderie and conformity. Thus, one is faced with the following dilemma: A vital and sustainable democracy requires the engagement of youth,

---

3.    See for instance the following summary of empirical evidence accumulated after Portugal de-criminalized the possession of all drugs in personal use: Transform (2014).

but youth politics acts negatively on the democratic vitality and sustainability. Accordingly, it might be regarded as ethically imperative to promote political engagement amongst youth, but to promote youth politics must be regarded as unethical. I will now present some empirical indicators from Norway, highlighting the problems.

## III. NORWAY AS CASE STUDY.

In this section, I will present empirical indicators from Norway.[4] With regards to Norwegians in general, only a small fraction have been involved in youth politics (ca 3%)[5], implying that the effects on voters are negligible, interesting only from a theoretical perspective. With regards to the candidates, however, the story is quite another.

There is a massive overrepresentation of ministers with a background in youth politics (ca 63%) compared with the population in general. From this point on, I will therefore focus on the negative effects of youth politics on the set of electable candidates. In section one, I argued that one way of increasing DVS is by increasing the suitability of the candidates that constitute the set of electable candidates. And it seems unreasonable to assume that Norway cannot achieve a more suitable set of electable candidates. That is to say, it seems unreasonable to assume that the most suitable candidates signal *de facto* superiority by engaging in youth politics.

One can therefore distinguish between two negative effects of youth politics. First, the direct effect: Elected candidates will be influenced negatively by camaraderie and conformity, as explained in the previous section. Second, the indirect effect: There appears to be some undesirable selection barriers which disincentivize potentially more suitable candidates to partake in politics. An example of such a barrier might be the network required to rise in the ranks; people that have not engaged in youth politics will be at a substantial disadvantage compared to those who have. How can one resolve these problems? That is what I will address in the final section.

---

4.    I characterize the findings as empirical 'indicators' as the data set is too small to amount to that which is commonly understood by empirical 'evidence.'

5.    The data on 'Norwegians in general' is extracted from The Norwegian Directorate for Children, Youth and Family Affairs (2015).

## IV. POTENTIAL SOLUTIONS TO THE DEMOCRAT-
## IC DILEMMA AND CONCLUDING REMARKS.

One thing is clear: The current celebration of youth politics must be regarded as unethical, as it enforces negative effects on DVS. This implies that society in general needs to make youth politics less attractive, for instance by paying less attention to the opinions of youth politicians. Other things are less clear, like what policies can be implemented in order to reduce youth politics? The most obvious policy would be to forbid youth from organizing politically. This would naturally be highly objectionable, as it gravely violates fundamental democratic principles such as freedom of speech and organization.

It seems like camaraderie and conformity become problems mainly because there is a career continuum from 'engaged in youth politics' to 'elected minister.' And in order to move along this continuum—that is, for the career-minded youth to rise through the ranks—negative effects kick in. A solution needs to create some sort of discontinuum. This may be done by setting a minimum age limit to holding any sort of public office. One might also require non-party work experience for eligibility in order to create a similar effect. By creating such discontinuums, one can potentially counter both the direct and indirect negative effects of youth politics. First, one might break down some of the deep relational commitments the members have to their political communities. Second, one might break down some of the barriers preventing potentially better candidates from getting involved with politics, that is candidates who have not been involved in youth politics.

It goes without saying that the effect of youth politics on democratic vitality and sustainability must be explored by further empirical research and philosophical inquiry. So must potential solutions. This paper has sought to initiate such research and discussion by challenging the collective celebration of youth politics.

### REFERENCES

Bureau for Development Policy. (2014). *Fast Facts: United Nations Development Programme*. United Nations Development Programme (UNDP). *United Nations.*

Ødegård, G., & Bergh, J. (2011). Unge i tradisjonell politikk. Institutt for samfunnsforskning: Deltakelse i valg, kommunestyrer og partier, Norway, p. 15.

Transform. (2014). *Drug decriminalisation in Portugal: setting the record straight*. Extracted from: http://www.tdpf.org.uk/resources/publications/drug-decriminalisation-portugal-setting-record-straight [Accessed November 2017]

The Norwegian Directorate for Children, Youth and Family Affairs: Barne-, Ungdoms- og familiedirektoratet. (2015). *Medlemskap i politiske ungdomspartier*. Extracted from: https://www.bufdir.no/Statistikk_og_analyse/Oppvekst/Samfunnsdeltakelse/Medlemskap_i_politiske_ungdomspartier/ [Accessed November 2017].